

**VŠB - Technická univerzita Ostrava**  
**Fakulta elektrotechniky a informatiky**

**BAKALÁŘSKÁ PRÁCE**

**2010**

**Hana Vašíčková**

**VŠB - Technická univerzita Ostrava**  
**Fakulta elektrotechniky a informatiky**  
**Katedra aplikované matematiky**

**Verifikace modelu toxicity na aktuálních  
onkologických datech**  
**Verification of toxicity models applied to  
current oncology data**

**2010**

**Hana Vašíčková**

VŠB - Technická univerzita Ostrava  
Fakulta elektrotechniky a informatiky  
Katedra aplikované matematiky

## Zadání bakalářské práce

Student: **Hana Vašíčková**  
Studijní program: B2647 Informační a komunikační technologie  
Studijní obor: 1103R031 Výpočetní matematika  
Téma: Verifikace modelu toxicity na aktuálních onkologických datech  
Verification of toxicity models applied to current oncology data

### Zásady pro vypracování:

Téma práce vzniklo na základě aktivní spolupráce s onkology z FN Ostrava Poruba. Modelování toxicity onkologických pacientů se doposud provádí na základě převzatých matematických modelů. Cílem této práce je aplikovat obecné principy těchto matematických modelů na aktuální data pacientů z FN a zejména posoudit jejich vhodnost a účelnost při jejich používání, případně výhody a nevýhody, to vše na základě aktivní kooperace s ošetřujícími lékaři. Postup práce:

1. Studium základů matematického modelování onkologických dat.
2. Organizace a zpracování dodaných onkologických dat.
3. Studium statistických testů pro posouzení významnosti koeficientů v použitých modelech, či modelů samotných.
4. Algoritmické zpracování a počítačová implementace těchto testů.
5. Provedení a vyhodnocení těchto testů na vzorku lékařských dat dle instrukcí vedoucího práce.
6. Vyhodnocení verifikace modelů.

### Seznam doporučené odborné literatury:

- R.Briš, M.Litschmannová, Statistika I, elektronické skriptum VŠB TUO, FEI, 2004.
- Hebák P., Hustopecký J., Malá I., Vícerozměrné statistické metody (2), INFORATORIUM 2005, ISBN 80-7333-036-9

Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí bakalářské práce: **doc. Ing. Radim Briš, CSc.**

Datum zadání: 20.11.2009

Datum odevzdání: 07.05.2010

*Z. Dostál*

prof. RNDr. Zdeněk Dostál, DSc.  
vedoucí katedry



*I. Vondrák*

prof. Ing. Ivo Vondrák, CSc.  
děkan fakulty

# Prohlášení

Prohlašuji, že jsem tuto bakalářskou práci vypracovala samostatně. Uvedla jsem všechny literární prameny a publikace, ze kterých jsem čerpala.

V Ostravě dne 15.4.2010

.....

# Poděkování

Děkuji tímto svému vedoucímu bakalářské práce Doc. Ing. Radimu Brišovi, Csc. za odbornou pomoc a konzultaci při vypracování, lékařům z Fakultní nemocnice Ostrava za poskytnutí dat ke zpracování a v neposlední řadě své rodině a přátelům za podporu.

# **Abstrakt a klíčová slova**

## **Abstrakt**

Tato práce se zabývá zpracováním lékařských onkologických dat z Fakultní nemocnice Ostrava. Výchozí data byla získána při léčbě karcinomu prostaty. Cílem je zodpovězení jednotlivých otázek, které byly položeny samotnými lékaři. Postup je založen na exploratorní analýze a kontingenčních tabulkách popisující jednotlivé závislosti. Závěr je zaměřen na křivky přežití popsané Kaplan-Meierovou metodou. Při řešení byly využity aplikace MS Excel 2007, SSPS 18.0 a Statgraphics plus 5.1.

## **Klíčová slova**

karcinom prostaty, exploratorní analýza, kontingenční tabulka, křivky přežití, Kaplan-Meier

# **Abstract and key words**

## **Abstract**

This project is focused on processing of carcinological data gained from Faculty Hospital of Ostrava. The data were obtained by therapy of prostate cancer. The goal is making out questions asked by physicians themselves. Process is based on exploratory analysis and contingency tables that describes particular dependencies. Final part contains Kaplan-Meier survival curves. There are used applications like MS Excel 2007, SSPS 18.0 and Statgraphics plus 5.1 by elaboration.

## **Key words**

prostate cancer, exploratory analysis, contingency table, survival curves, Kaplan-Meier

# Seznam použitých symbolů a zkratek

<b>PSA</b>	Prostatický specifický antigen, koncentrace enzymu v krvi, který určuje aktivitu onemocnění, čím je vyšší, tím je onemocnění pokročilejší
<b>CT</b>	Computer tomograf
<b>TNM</b>	klasifikace, která určuje rozsah postižení nádorovým onemocněním
<b>FNOP</b>	Fakultní nemocnice Ostrava Poruba
<b>RAPE</b>	radikální prostatektomie - chirurgické vyjmutí prostaty
<b>LAE</b>	pánevní lymfadenektomie - odstraněním pánevních lymfatických uzlin
<b>OE</b>	orchiektomie – chirurgické odstranění varlat
<b>K</b>	kurativní typ radioterapie - je ambice pacienta vyléčit i za cenu nežádoucích účinků
<b>P</b>	paliativní typ radioterapie - stav je nevyléčitelný, léčba je vedena šetrně s minimem nežádoucích účinků
<b>RTOG</b>	Radiation Therapy Oncology Group - stupnice pro hodnocení chronické radiační toxicity
<b>UGS</b>	urogenitální soustava - intenzita nežádoucích účinků radioterapie v močovém měchýři
<b>GIT</b>	gastrointestinální soustava - intenzita nežádoucích účinků radioterapie v konečníku
<b>GS</b>	(Gleason score) popisuje agresivitu nádoru na základě vyšetření pod mikroskopem

# Seznam obrázků

Obrázek 1: Nesjednocený formát .....	18
Obrázek 2: Nesjednocené označení .....	18
Obrázek 3: Stejný význam různých označení.....	19
Obrázek 4: Klasifikace pacientů do tříd .....	19
Obrázek 5: Schéma typu proměnných v exploratorní analýze .....	22
Obrázek 6: Mozaikový graf.....	24
Obrázek 7: Kumulativní sloupcový graf .....	24
Obrázek 8: Shlukový graf.....	25
Obrázek 9: Graf pacientů po a bez RAPE .....	32
Obrázek 10: Graf radioterapie v jednotlivých stádiích.....	33
Obrázek 11: Graf radioterapie typu K v jednotlivých stádiích.....	34
Obrázek 12: Graf závislosti typu operace ve skupinách.....	35
Obrázek 13: Graf typu radioterapie po jednotlivých operacích.....	37
Obrázek 14: Graf typu radioterapie po operaci RAPE .....	37
Obrázek 15: Výšečový graf typu radioterapie po operaci RAPE .....	38
Obrázek 16: Graf použití hormonů ve skupinách.....	39
Obrázek 17: Graf použití hormonů v jednotlivých stádiích .....	40
Obrázek 18: Graf akutní UGS bez RAPE a po RAPE.....	41
Obrázek 19: Graf akutní GIT bez RAPE a po RAPE.....	42
Obrázek 20: Graf pozdní UGS bez RAPE a po RAPE.....	43
Obrázek 21: Graf pozdní GIT bez RAPE a po RAPE .....	44
Obrázek 22: Graf akutní UGS bez RAPE .....	45
Obrázek 23: Graf akutní GIT bez RAPE .....	46
Obrázek 24: Graf pozdní UGS bez RAPE.....	47
Obrázek 25: Graf pozdní GIT bez RAPE .....	48
Obrázek 26: Celková křivka přežití.....	49
Obrázek 27: Křivky přežití pro M0 a M1 .....	50
Obrázek 28: Křivky přežití stádií M0.....	52
Obrázek 29: Křivky přežití M0 se závislostí na hormonech .....	54

# Seznam tabulek

Tabulka 1: Schéma kontingenční tabulky .....	23
Tabulka 2: Závislost rizikové skupiny na typu radioterapie.....	33
Tabulka 3: Závislost rizikové skupiny na typu operace .....	35
Tabulka 4: Závislost typu radioterapie na typu operace.....	38
Tabulka 5: Závislost podání hormonů na rizikové skupině.....	39
Tabulka 6: Závislost akutní UGS na podstoupení RAPE.....	41
Tabulka 7: Závislost akutní GIT na podstoupení RAPE .....	42
Tabulka 8: Závislost pozdní UGS na podstoupení RAPE .....	43
Tabulka 9: Závislost pozdní GIT na podstoupení RAPE .....	44
Tabulka 10: Závislost akutní UGS bez RAPE na stádiích M0 .....	45
Tabulka 11: Závislost akutní GIT bez RAPE na stádiích M0 .....	46
Tabulka 12: Závislost pozdní UGS bez RAPE na stádiích M0 .....	47
Tabulka 13: Závislost pozdní GIT bez RAPE na stádiích M0 .....	48
Tabulka 14: Shrnutí pro celkovou křivku přežití.....	49
Tabulka 15: Tabulka středních hodnot pro celkovou křivku přežití.....	50
Tabulka 16: Tabulka percentilů pro celkovou křivku přežití .....	50
Tabulka 17: Shrnutí křivek přežití pro M0 a M1 .....	51
Tabulka 18: Tabulka středních hodnot křivek přežití pro M0 a M1.....	51
Tabulka 19: Tabulka percentilů křivek přežití pro M0 a M1 .....	51
Tabulka 20: Shrnutí křivek přežití pro stádia M0.....	52
Tabulka 21: Tabulka středních hodnot pro křivky přežití stádií M0 .....	53
Tabulka 22: Tabulka percentilů křivek přežití pro stádia M0 .....	53
Tabulka 23: Shrnutí křivek přežití pro závislost skupiny M0 na hormonech.....	54



# Obsah

1. Úvod .....	11
2. Rakovina prostaty.....	12
2.1. Popis rakoviny prostaty .....	12
2.1. Etiologie, rizikové faktory a prevence rakoviny prostaty .....	12
2.2. Příznaky a průběh rakoviny prostaty.....	13
2.3. Druhy vyšetření .....	13
2.4. Stupně a stádia rakoviny .....	14
2.5. Léčba rakoviny prostaty .....	14
3. Poznámky k analýze medicínských dat.....	16
3.1. Spolupráce lékařů a statistiků .....	16
3.2. Hledání společného jazyka .....	16
3.3. Pozice statistika ve výzkumu .....	16
3.4. Formulace hypotézy a doporučení designu .....	16
3.5. Volba statistického modelu (testu) .....	17
3.6. Odhad rozsahu výběru a nastavení randomizace .....	17
3.7. Organizace sběru dat.....	17
4. Exploratorní analýza proměnných .....	21
5. Kontingenční tabulka a její testy .....	23
Schéma kontingenční tabulky.....	23
Testy v kontingenční tabulce.....	25
$\chi^2$ test nezávislosti v kontingenční tabulce .....	25
6. Cenzorovaná data.....	27
Výběrové plány.....	27
Cenzorování typu I (cenzorování časem) .....	27
Cenzorování typu II (cenzorování poruchou).....	27
Náhodné cenzorování .....	27
7. Neparametrické metody .....	29

Kaplan-Meierův odhad (angl. Product limit estimate) .....	30
8. Vlastní zpracování dat .....	32
Jednotlivé řešené otázky .....	32
1. Kolik pacientů v jednotlivých stádiích podstoupilo radioterapii –K resp. ....	33
2. Kolik pacientů v jednotlivých stádiích podstoupilo RAPE či LAE či kombinaci, OE.....	35
3. Kolik pacientů po RAPE mělo radioterapii K-A, kolik K-S.....	37
4. Kolik pacientů v jednotlivých stádiích mělo nějaký typ hormonální léčby.....	39
5. Jaké je procentuální zastoupení jednotlivých stupňů akutní a pozdní toxicity (akutní RTOG a pozdní RTOG) po RAPE a bez RAPE .....	41
6. Zpracování křivek přežití .....	49
9. Závěr .....	55
Literatura .....	56
Přílohy.....	57

# 1. Úvod

Tato práce se zabývá zpracováním lékařských onkologických dat z Fakultní nemocnice Ostrava. Výchozí data byla získána při léčbě karcinomu prostaty, který patří mezi nejčastější zhoubné nádory. Cílem této práce je zodpovězení jednotlivých otázek, které byly položeny samotnými lékaři. Práce zahrnovala časté konzultace s danými lékaři o nejrůznějších podrobnostech ohledně databáze, ze které byla data vybrána. Problémy s databází jsou shrnuty v části Poznámky k analýze medicínských dat. Tenhle článek byl zároveň prezentován na konferenci v Liberci, které jsme se zúčastnili.

V první části je práce zaměřena na teorii potřebnou ke zpracování zadaných otázek. Zabývá se zde základními informacemi o karcinomu prostaty, exploratorní analýzou, kontingenčními tabulkami a křivkami přežití. V následující části pak s použitím těchto metod za pomoci softwaru MS Excel 2007, SSPS 18.0 a Statgraphics plus 5.1 jsou otázky zodpovězeny.

Cílem této práce je seznámit se s praktickým využitím statistiky na medicínských datech, vyzkoušet si spolupráci s lékaři na konkrétních problémech a porozumět jejich terminologii.

## 2. Rakovina prostaty

### 2.1. Popis rakoviny prostaty

Karcinom prostaty tvoří asi 4% všech nádorových onemocnění. V České republice se jeho incidence pohybuje okolo 60 případů na 100 000 mužů. V mužské populaci tedy představuje tato rakovina 10% všech nádorů podobně jako v jiných zemích západní Evropy. Výskyt latentního onemocnění (tzn. skrytého onemocnění, které se ještě nijak neprojeví a nepřivede proto pacienta k lékaři, nemůže být tudíž ani součástí epidemiologických studií) je však mnohem vyšší.

S touto rakovinou se lékaři setkávají stále častěji. V České republice stoupla incidence až o 70% oproti počátku devadesátých let. Nejvíce nemocných je ve věku **mezi 70. – 80. rokem**, přibývá ale i pacientů mladších, proto se uvažuje i o **screeningu**, který by mohl zachytit nádor třeba u šedesátiletého pacienta. Nevýhodou by byla docela zbytečná zátěž pro starší pacienty, kteří již mají další, třeba také nádorová, onemocnění vážnějšího charakteru a karcinom prostaty tak u nich ustupuje do pozadí.

### 2.1. Etiologie, rizikové faktory a prevence rakoviny prostaty

Riziko karcinomu prostaty stoupá s **věkem**. Mladší muži (pod 45 let) tvoří jen jeden případ na 100 000 mužů, ale muži ve věku nad 80 let představují již 735 případů na 100 000 mužů. Znamená to, že rakovina prostaty je pro starší muže naprosto běžná a pokud se dožijí 90ti let, budou nádor s největší pravděpodobností mít, ačkoliv se u nich nemusí nijak projevit a oni se o něm tedy ani nedozví.

Dalším rizikovým faktorem je **rasa**. Rakovina prostaty je skoro dvakrát častější u **černochů** a je zde také vyšší tendence k výskytu v mladším věku.

Roli hrají samozřejmě i **genetické faktory**. Přímé příbuzenství prvního stupně s nemocným karcinomem prostaty (tedy syn nemocného otce) znamená poněkud vyšší riziko než v běžné populaci.

Předpokládá se, že jídelníček s **vysokým obsahem tuků** a kadmium obsažené v **cigaretovém kouři** mohou mít vztah k etiologii (etiologie = původ onemocnění) této rakoviny. Nejdůležitější prevencí je tudíž zdravý životní styl – vhodná úprava masa, omezení živočišných tuků, preferování polyenových nenasycených mastných kyselin (omega 3), dostatek ovoce, zeleniny a tím i vitamínů, které mají antioxidační efekt (vit. C, E) a nekuřáctví.

Součástí zdravého života by měl být také každodenní pohyb a omezení nadměrné konzumace alkoholu.

Touto komplexní prevencí se předchází nejen rakovině prostaty, ale i dalším mnohem závažnějším rakovinám a jiným onemocněním. Například vynecháním kouření a tučného jídla, se mnohonásobně sníží riziko kardiovaskulárního onemocnění (např. infarkt myokardu).

Nadbytek alkoholu je škodlivý pro játra (cirhóza jater) a zároveň významně přispívá ke vzniku některých nádorů (především v oblasti krku, úst a jazyka). Vhodným složením potravy se ovlivňuje výskyt nádorů trávicího traktu, nekuřáctvím se snižuje výskyt karcinomu plic, ačkoliv tam se tento příznivý efekt projeví až o několik let po tom, když člověk přestane kouřit. Je třeba si uvědomit, že v cigaretách je obsaženo několik desítek i set karcinogenních látek, které mohou způsobit vznik

nádorového bujení a také ničí výstelku stěny cév, kde vzniká na tomto podkladě ateroskleróza se všemi možnými důsledky (infarkt myokardu, angina pectoris, mrtvice, ischemická choroba dolních končetin a další).

Mezi nejednoznačné rizikové faktory patří **sexuálně přenosná infekční onemocnění** a s tím související míra sexuální aktivity. Ačkoliv *testosteron* – mužský pohlavní hormon – nemá přímý vztah ke vzniku nádoru, může se podílet na rozvoji klinické manifestace. Bylo zjištěno, že incidence karcinomu prostaty je mnohem nižší u eunuchů – ti postrádají testosteron.

Diskutovanou povinnou prevencí je již výše zmíněný **screening**. Mohl by se týkat zatím jen rizikových skupin, které však nejsou ještě jednoznačně definované. Vyšetřoval by se jeden krevní marker – **prostatický specifický antigen (PSA)** – který je charakteristický právě pro karcinom prostaty (vylučuje se do krve díky nádorovým buňkám prostaty), v séru zdravých jedinců nebo jinak nemocných než s karcinomem prostaty ho nenajdeme.

Kromě toho je doporučováno **vyšetření per rectum** jednou ročně u mužů nad čtyřicet let.

## 2.2. Příznaky a průběh rakoviny prostaty

Časná stadia karcinomu prostaty jsou většinou asymptomatická (tzn. bez příznaků) a mohou být zachycena v některých případech během operace močové trubice z jiných důvodů.

U lokálně pokročilých nádorů (to jsou takové, které už se šíří do okolí – do močového měchýře, do močové trubice, ale ještě nemetastazovaly do vzdálených orgánů) jsou hlavními příznaky pacientů různé **potíže při močení**. Pacienti chodí na toaletu v kratších intervalech a také častěji v noci nebo mohou naopak moč zadržovat a vymočení je pro ně náročné.

Celkem vzácné je objevení **krve v moči** a onemocnění ledvin (ledvinná nedostatečnost, resp. selhávání ledvin).

Známkou velmi pokročilých nádorů pak bývá nejčastěji **bolest** z kostního metastatického procesu, **hlavně zad, kyčlí a končetin**. Metastáza v kostech může někdy utlačovat míchu a působit potíže s hybností.

Celkovými příznaky jako u jiných nádorových onemocnění je **nechutenství, hubnutí, únava** a celková **slabost**.

Nádory prostaty se mohou mezi sebou histologicky lišit, některé rostou velmi pomalu a nemetastazují, takže mohou mít vcelku dobrou prognózu, ačkoliv způsobují třeba lokální potíže s močením.

Nádory, které rostou rychle a metastazují, jsou naopak nepříznivé. Metastázy se nacházejí nejčastěji v kostech, dále v plicích, játrech, méně v nadledvinách a ledvinách.

## 2.3. Druhy vyšetření

**Rektální vyšetření (per rectum)** - lékař vkládá namazaný prst v rukavici do konečníku, aby mohl prohmatat zadní stěnu žlázy prostaty. Tímto vyšetřením lékař zjišťuje, zda nejsou na prostatě žádné nepoddajné oblasti, tvrdé uzlíky nebo jiné nepravidelnosti.

**Vyšetření PSA** - je to test na enzym produkovaný buňkami prostaty, jak normálními, tak rakovinovými. Je-li prostata poškozená, uniká do krevního oběhu vyšší množství PSA. Zvýšené PSA však nemusí znamenat ještě rakovinu, ale také např. zánět močových cest. Nicméně čím vyšší je hladina PSA, tím je vyšší pravděpodobnost přítomnosti rakoviny. PSA se také využívá jako

monitoring léčby, zda léčba funguje a zda se rakovina nevrací. Každá léčba očekává jinou hladinu PSA.

Pokud lékař rakovinu zjistí, následují další vyšetření, která se dělají pro správnou diagnózu a určení léčby. Lékař také potřebuje vědět, zda se rakovina nerozšířila ven z prostaty.

**Scan skeletu** - nukleární lékařská technika. Při tomto vyšetření jsou malé kousky radioaktivní látky injektovány do krevního oběhu, aby se pak zobrazil Váš skelet a abnormality na něm. Scan nezobrazuje rakovinu, ale ukazuje oblasti rychlého růstu kostí. Pro rakovinu prostaty je typické, že se na kostru rozšiřuje obvykle ve formě náhodných horkých skvrn. K rakovinám rozšiřujících se do kostí patří nejčastěji rakovina prsu, prostaty, močového měchýře a tlustého střeva.

**CT vyšetření** - počítačová axiální tomografie - ukáže lékaři vnitřní orgány.

## 2.4. Stupně a stádia rakoviny

K popisu rozsahu malignity (zhoubnosti) rakoviny se používají stupně, které slouží jako standardní měřítko. Některé rakoviny jsou velmi agresivní (buňky se změní tak, že se nepodobají vlastním tkáním) - takové jsou charakterizovány vysokým stupněm. Jiné buňky se mohou podobat normálním buňkám s malými výjimkami - tyto buňky se popisují nízkým stupněm. Pro stupňování se používá **GLEASONŮV SYSTÉM** (měřítko). Váš lékař pak může mluvit o rakovině jako o Gleasonu 7 nebo jednoduše o stupni 5. Přičemž rakovina nízkého stupně je méně nebezpečná (při použití Gleasonova systému by to bylo 2,3 nebo 4) a naopak (vysoký stupeň 8,9 a 10).

Stádia rakoviny popisují, je-li rakovina malá a zůstává v prostatě nebo je-li velká a rozšířila se na jiné orgány.

**Stadium A:** Rakovina zjištěná náhodně nebo na základě zvýšeného PSA.

**Stadium B:** Rakovina zjištěná na základě digitálního rektálního vyšetření - zůstává v prostatě.

**Stadium C:** Rakovina rozšířená na tkáň vně prostaty.

**Stadium D:** Rakovina rozšířená na lymfatické uzliny nebo kosti.

Pro určení stádia rakoviny se častěji používá **Mezinárodní systém TNM**, kdy jednotlivá písmena znamenají:

**T:** Popisuje samotnou rakovinu číselným označením udávajícím velikost

**N:** Popisuje rozšíření na lymfatické uzliny

**M:** Informuje, zda se rakovina rozšířila a stala se metastatickou

Podle tohoto označení by Vaše diagnóza mohla znít T2N0M0.

Nádory prostaty rozdělujeme do 2 základních skupin: M0 a M1. Skupinu M0 - rozdělujeme do 3 stádií dle rizika (nízké, střední, vysoké)

Nízké: < T2 a zároveň GS < 7 a zároveň PSA <10

Střední: T2 nebo GS 7, 10 < PSA < 20

Vysoké: ostatní

## 2.5. Léčba rakoviny prostaty

Způsob léčby závisí na povaze nádoru, věku nemocného, jeho celkovém stavu a přidružených jiných vážnějších onemocnění zvláště ve vyšším věku.

Nemocní, u nichž se na karcinom přišlo náhodou během jiné operace v této oblasti a je tedy teprve v počátečním stádiu, jsou pouze sledováni. Lokálně pokročilá stadia (nádor už se rozšířil do uzlin a do močové trubice) jsou vhodná k **operaci**.

Nejzávažnější komplikací může být **inkontinence** – nemožnost zadržení moče v močovém měchýři spojená s pomočováním – a **impotence**. Dnes už však existují dokonalejší operační techniky, takže k těmto komplikacím tak často nedochází.

Další léčebnou možností je **ozařování (radioterapie)** oblasti prostaty nebo celé malé pánve a v místech případného metastatického postižení (především kostí). Na některých pracovištích se dělá i **brachyradioterapie** (ozařování „zevnitř“, kdy se pomocí speciálních instrumentů zavede zářič do blízkosti prostaty, paprsky tedy nejdou přes kůži, nemají tolik vedlejších účinků a jejich léčebný efekt je vyšší). Výsledky ozařování nádoru jsou srovnatelné s chirurgickou léčbou.

Výjimečností nádoru prostaty je, že se v jeho léčbě dá použít i **hormonální terapie** (podobně jako u nádoru prsu u žen). Dříve se užívaly **estrogeny** (ženské pohlavní hormony, které mají na karcinom prostaty tlumivý efekt, protože inhibují uvolňování jednoho hormonu hypofýzy – luteinizačního hormonu – a díky jeho nedostatku klesá tvorba testosteronu a snižuje se růst buněk prostaty), ale po tom, co se ukázaly jejich vedlejší nežádoucí účinky, zejména tromboembolické a kardiovaskulární komplikace, se používat prakticky přestaly.

Jinou, v dnešní době preferovanou, možností hormonálního zásahu je **operace varlat – jejich vyjmutí (orchiektomie)**. Tato operace je velmi jednoduchým chirurgickým výkonem. Přestane se tvořit testosteron (produkují ho totiž buňky pohlavních mužských žláz – varlat), který ovlivňuje růst nádoru.

Analogií je podávání **gonadoliberinů**, resp. jejich náhražek (goserelin, leuprolid a další), které mají právě i opačný účinek, než přirozené hormony. Dochází tedy také k inhibici produkce testosteronu v podstatě jako při podávání estrogenů s tím rozdílem, že tady to máme bez vedlejších účinků.

Z psychologického hlediska je tato léčba lépe akceptována než chirurgická kastrace, ale je také významně dražší. Na počátku léčby může dojít k přechodnému zhoršení, což se může projevit například intenzivnějšími bolestmi kostních metastáz. Prevencí proti tomuto fenoménu je současné podání **antiandrogenů** (látky, která má účinek opačný než testosteron) zpravidla již týden před plánovaným zahájením léčby. K výraznému snížení hladin testosteronu (stejných jako kdyby byla provedena operační kastrace) dojde po třech týdnech léčby.

Jindy se antiandrogeny podávají současně celou dobu s gonadoliberinovými analogy, hovoříme pak o **kombinované léčbě**. Jde o to, že testosteron se zčásti tvoří i v nadledvinách, proto i po zastavení jeho tvorby ve varleti mohou jeho, sice už nižší, hladiny způsobovat nadále růst nádoru prostaty.

Pokročilá stadia nádoru, pokud na něj nezabrala hormonální léčba, je možné léčit **chemoterapií**. Vždy je to ale až druhořadý postup po té, co je vyzkoušena léčba hormonální. Podává se mitoxantron, docetaxel a vinorelbin. Estramustin je zvláštní druh léku, jehož efekty zatím nejsou prokazatelné.

## 3. Poznámky k analýze medicínských dat

Tento článek byl prezentován na konferenci Request '09 (26.11.2009 - 27.11.2009) v Liberci a následně vydán ve sborníku přednášek. Obsahuje nedostatky databáze, se kterými jsem se musela během práce potýkat.

### 3.1. Spolupráce lékařů a statistiků

V dnešní době má jen málokdo možnost věnovat se studiu oborů hraničící s jeho vlastní profesí. Na katedře aplikované matematiky, FEI, VŠB-TU Ostrava se, zejména v souvislosti se spoluprací s FNOP, setkáváme s požadavky na analýzu medicínských dat. Práce je to velmi zajímavá – jde o oblast, která se určitým způsobem dotýká každého z nás. Lékaři řeší spoustu zajímavých, často i složitých problémů, které vedou na více či méně sofistikované statistické metody a modely. Snažíme se data analyzovat a interpretovat zjištěné závěry společně s lékaři, a tím částečně přispívat ke zkvalitnění zdravotní péče.

### 3.2. Hledání společného jazyka

Aby byla možná spolupráce mezi lékaři a statistiky, musíme i my nahlédnout do tajemství medicíny a lékaři musí být obeznámeni se základy statistiky, abychom při řešení konkrétního problému našli společný jazyk. Podaří-li se nám nastínit lékařům způsob našeho uvažování, komunikace se stává jednodušší a efektivnější. Cílem této spolupráce je zajistit správnost a odbornost lékařského výzkumu po stránce statistického vyhodnocování dat a interpretace získaných výsledků.

### 3.3. Pozice statistika ve výzkumu

Jedním z nedostatků naší spolupráce je to, že jsme často považováni pouze za kvalifikované „zpracovatele dat“ – dat, která někdo jiný nashromáždí. Samozřejmě, že i toto je naše práce, ale pak by po nás nemělo být požadováno zobecňování zjištěných údajů – odhad průměrné roční spotřeby určitého typu léčiv v nemocnici, srovnání relativních četností výskytu onemocnění u mužů a žen, apod.. Exploratorní statistika je postačující statistickou metodou pouze v případech, kdy výstupem analýzy je informace o situaci v daném čase. Ve všech ostatních případech je nutné, aby spolupráce lékaře se statistikem začala již ve fázi plánování projektu. Statistik by se měl podílet na navrhování experimentů a studií, resp. na navrhování a vytváření vhodných dotazníků pro sběr dat v rámci epidemiologických studií (položková analýza a vývoj testu).

### 3.4. Formulace hypotézy a doporučení designu

V první fázi by měl statistik na základě znalosti primárního cíle studie pomoci zformulovat a jasně definovat hypotézu, resp. hypotézy, které by studie měla ověřit.

Následuje doporučení designu studie. Standardem v klinických studiích jsou randomizovaná paralelní studie (s experimentálním a kontrolním ramenem) a cross-over (zkřížená) studie. Ze statistického hlediska je výhodnější cross-over studie, která umožňuje lepší odhad individuální odpovědi subjektu (pacienta). Musíme však předem zvážit, zda je tento design možný. Lze u pacienta provést „zkřížený“ experiment? Nemůže být pacient již v první fázi experimentu vyléčen, resp. stabilizován? Vzáli jsme v úvahu všechny etické důvody? Jsme si vědomi větší časové náročnosti cross-over studie?



Nezapomeňme ani na možnost studií bez intervence (bez terapeutického zásahu) – studie case control (Co je příčinou toho, že u někoho se nemoc vyvinula a u někoho ne?, retrospektivní design), kohortová studie (Máme výběr z populace, která je „v možném riziku“. Časová náročnost studie, cenzorování...), průřezová studie (Co se děje právě teď?)

### **3.5. Volba statistického modelu (testu)**

Máme-li již jasně definovanou hypotézu (soubor hypotéz) i populaci subjektů, pro něž má hypotéza smysl, stojíme před otázkou jak hypotézu ověřit (testovat). K tomu je třeba vybrat statistický test (model), který bude použit a stanovit hladinu významnosti, na níž budeme rozhodovat. Popřípadě tento model zvolit kondicionálně, dle nějaké, předem neznámé, vlastnosti dat. Volíme-li test tak, aby nám vyšlo to, co potřebujeme, neměli bychom mluvit o statistice.

### **3.6. Odhad rozsahu výběru a nastavení randomizace**

Dalším krokem by měl být odhad počtu pacientů potřebných pro studii (sample size determination). V případě, že zvolený test umožňuje tento odhad (t-test, test relativních četností), může tento krok zamezit tomu, že klinicky relevantní rozdíl nebude vyhodnocen jako statistický významný (malý rozsah výběru, underpowered studie), popř. že za statisticky významný bude označen klinicky irelevantní rozdíl (velký rozsah výběru, overpowered studie).

Samotnému sběru dat by mělo ještě předcházet nastavení randomizace, jejímž cílem je zamezit subjektivnímu a selektivnímu rozdělování pacientů do ramen studie. V praxi se nejčastěji setkáváme s těmito typy randomizace:

- Randomizace podle pořadového čísla vstupu pacienta do studie, resp. podle iniciál pacienta
- Kompletní randomizace
- Bloková randomizace
- Stratifikační permutační bloková randomizace
- Adaptivní randomizace

### **3.7. Organizace sběru dat**

Při samotném sběru dat je třeba mít na mysli, že data se jednou budou statisticky zpracovávat. I v této fázi návrhu experimentu je vhodná spoluúčast statistika. V případě, že lékaři nemají k dispozici profesionální software pro zápis sledovaných proměnných pro jednotlivé subjekty, používají nejčastěji MS Excel. Uvedeme několik příkladů z nevhodně sestavené databáze v MS Excel.

- Nesjednocený formát

	A	B	C	D	E	F	G	H	I	J	K
1	Th:	Operace:	Radioterapie:								
2		(RAPE, LAE, RAPE+LAE, OE)	datum	Kurativní (K, A, adj., K+S-salvage)	Paliativní (P, P+SIV, pánev)	(0-10mm)	(3D, IMRT, SBRT, ART)	datum	datum	akutní RTOG (0-4)	
3	Poř.č.	typ	dne	typ	PTV1	lem CTV, PTV	RT	od	do	UGS	GIT
4	1	0	0	K	1	10	3D	5.1.2006	28.2.2006	0	0
5	2	0	0	K	1	10	3D	9.1.2006	1.3.2006	0	0
6	3	0	0	K	1	10	3D	23.1.2006	9.3.2006	0	0
7	4	0	0	K	2	10	3D	23.1.2006	17.3.2006	0	0
8	5	0	0	K	1	10	3D	25.1.2006	16.3.2006	1	0
9	6	0	0	K	3	10	3D	2.2.2006	27.3.2006	0	0
10	7	0	0	K	1	10	3D	31.1.2006	27.3.2006	1	1
11	8	0	0	K	1	10	3D	8.2.2006	3.4.2006	0	0
12	9	OE	24.1.2006	K	1	10	3D	8.2.2006	3.4.2006	0	1
13	10	RAPE	10.1.2006	K-A	1	10	3D	23.2.2006	18.4.2006	1	0
14	11	RAPE+LAE	18.1.2006	K-A	2	10	3D	27.2.2006	19.4.2006	1	1
15	12	0	0	K	3	10	3D	2.3.2006	25.4.2006	0	0
16	13	0	0	K	1	10	3D	1.3.2006	24.4.2006	0	1
17	14	0	0	K	1	10	3D	1.3.2006	24.4.2006	2	2
18	15	0	0	K	1	10	3D	6.3.2006	27.4.2006	1	1
19	16	0	0	K	1	10	3D	13.3.2006	15.5.2006	1	2
20	17	0	0	K	3	10	3D	7.3.2006	12.5.2006	1	1

Obrázek 1: Nesjednocený formát

- Nesjednocené označení

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
289	286	3	X	1	4	44,9	09/2006							
290	287	1c	0	0	nelze	8,06	3.10.2007	X						
291	288	3a	0	0	7	6,5	15.8.2007							
292	289	2b	0	0	X	4,64	15.4.2004							
293	290	2a	0	0	7	11	10.7.2007							
294	291	X	1	1	neuv.	1424	23.11.2007							
295	292	3b	1	0	8	13	15.5.2001							
296	293	3	2	0	6	23,9	14.1.2004						X	
297	294	1	0	0	4	1	27.11.2007						X	
298	295	2b	0	0	6	X	12.11.2003			X			X	
299	296	1c	0	0	neuv.	3,75	31.10.2007	X						
300	297	2a	0	0	5	7,11	29.11.2007							
301	298	2c	0	0	7	0,79	30.10.2007						X	
302	299	2a	0	0	5	4	31.10.2007			X				
303	300	2a	0	0	7	11,21	23.11.2007							
304	301	2b	0	0	7	14	17.12.2007		X					
305	302	2b	0	0	6	6	15.11.2007							
306	303	2c	0	0	7	34	21.9.2007		X		X		X	
307	304	3a	0	0	X	X	3.1.1996		X				X	
308	305	4	0	0	8	65	17.8.2007	X					X	
309	306	1c	0	0	5	14	13.11.2007							
310	307	3a	0	0	7	42,5	13.4.2007	X		X		X		
311	308	2b	0	0	7	4,72	30.11.2007							
312	309	2	0	0	6	23	25.6.2007							
313	310	1	0	0	6	6,8	11.11.2004	X		X				
314	311	3	X	1	10	680	2.3.2007							
315	312	1c	0	0	6	7,42	7.12.2007							
316	313	2	0	0	5	11	10.1.2008					X		
317	314	3b	0	0	9	84	22.1.2008							
318	315	1c	0	0	6	6,9	3.1.2008			X				
319	316	1c	0	0	5	4,7	17.1.2008				X	X		
320	317	4	X	1	X	4,9	21.1.2008							x

Obrázek 2: Nesjednocené označení

- Stejný význam různých označení

	A	B	C	D
121	118	0	0	K
122	119	OE	31.8.2005	P
123	120	RAPE+LAE	27.5.2005	K-S
124	121	OE	22.7.2003	P
125	122	OE	15.5.2006	P
126	123	RAPE	15.8.2001	K-S
127	124	OE	13.12.2004	K
128	125	OE	5.5.2004	P
129	126	OE	15.6.2006	P
130	127	RAPE+LAE	28.3.2003	K-S
131	128	RAPE+OE	2003	P
132	129	OE RAPE	1995, 1998	K-S
133	130	OE	11.5.2004	P
134	131	0	0	P
135	132	OE	18.8.2005	P
136	133	OE	27.9.2006	P
137	134	RAPE+LAE	4.10.2001	K-S
138	135	0	0	K
139	136	RAPE+LAE	16.10.2002	K-S
140	137	RAPE	19.9.2005	K-S
141	138	OE	9.1.2006	P
142	139	RAPE+LAE	3.9.2003	K-S
143	140	OE	23.8.2006	P
144	141	RAPE+LAE	18.9.2002	K-S
145	142	RAPE+LAE	5.3.2003	K-S
146	143	RAPE+LAE+OE	2005	K-S
147	144	OE	11.5.2006	P
148	145	RAPE+LAE	5.12.2000	K-S
149	146	RAPE+LAE	5.9.2003	K-S
150	147	0	0	K
151	148	RAPE+LAE	2.2.2004	K-S
152	149	OE	15.12.2006	K

	A	B	C	D
343	340	0	0	K
344	341	0	0	K
345	342	RAPE+LAE	21.1.2008	K-A
346	343	0	0	K
347	344	0	0	K
348	345	0	0	K
349	346	0	0	K
350	347	0	0	K
351	348	0	0	K
352	349	0	0	K
353	350	0	0	K
354	351	LAEaOE	29.2.a.11.3.2008	P
355	352	LAE, OE	04/08	P
356	353	0	0	K
357	354	RAPE+LAE	19.6.2007	K-S
358	355	0	0	K
359	356	RAPE+LAE	16.5.2003	K-S
360	357	OE	2.4.2008	P
361	358	0	0	K
362	359	LAE	23.4.2008	K-A
363	360	RAPE	6.1.2006	K-S
364	361	0	0	K
365	362	RAPE+LAE	6.2.2008	K-A
366	363	0	0	K
367	364	0	0	K
368	365	0	0	K
369	366	OE	6.5.2008	P
370	367	0	0	K
371	368	0	0	K
372	369	0	0	K
373	370	0	0	K
374	371	0	0	K

Obrázek 3: Stejný význam různých označení

- Klasifikace pacientů do tříd

Schránka	Písmo	Zarovnání	Číslo	Styl	Buňky	Úpravy																							
M4	=KDYŽ(NEBŮ(C4="";G4="";I4="";L4="");"";KDYŽ(L4="1";"M1";KDYŽ(A(C4=1;G4<7;I4<10);"NÍZKÉ RIZIKO";KDYŽ(NEBŮ(C4=3;G4>7;I4>20);																												
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y				
1	Dg:																												
2	(1a-c,2a-c,3c-b,4)																												
3																													
4	1	2	2	0	0	2	2	17	17	0	0		STADIA	hist.	další dg. (označit "x")														
5	2	2a	2	0	0	6	6	4,5	4,5	0	0		STŘEDNÍ RIZIKO	24.11.2005										1					
6	3	2a	2	0	0	5	5	5,6	5,6	0	0		STŘEDNÍ RIZIKO	10.11.2005			X												
7	4	2a	2	0	0	7	7	26,9	26,9	0	0		VYSOKÉ RIZIKO	18.11.2005			X												
8	5	2b	2	0	0	4	4	4,5	4,5	0	0		STŘEDNÍ RIZIKO	22.12.2005															
9	6	2b	2	0	0	6	6	13,9	13,9	0	0		STŘEDNÍ RIZIKO	6.1.2006					X										
10	7	1c	1	0	0	4	4	7	7	0	0		NÍZKÉ RIZIKO	4.11.2005															
11	8	1c	1	0	0	3	3	5,2	5,2	0	0		NÍZKÉ RIZIKO	4.1.2006															
12	9	2b	2	0	0	5	5	12,5	12,5	0	0		STŘEDNÍ RIZIKO	6.12.2005															
13	10	3a	3	0	0	5	5	4,1	4,1	0	0		STŘEDNÍ RIZIKO	1.12.2005	X		X	X				C20	27.5.1999	1					
14	11	3b	3	0	0	6	6	22,5	22,5	0	0		VYSOKÉ RIZIKO	20.10.2005	X														
15	12	3	3	0	0	2	2	167	167	0	0		VYSOKÉ RIZIKO	9.11.2005										1					
16	13	2a	2	0	0	2	2	6,13	6,13	0	0		STŘEDNÍ RIZIKO	20.12.2005															
17	14	2c	2	0	0	6	6	3,2	3,2	0	0		STŘEDNÍ RIZIKO	27.1.2006															
18	15	2a	2	0	0	5	5	3,2	3,2	0	0		STŘEDNÍ RIZIKO	2.12.2005	X	X													
19	16	2c	2	0	0	5	5	12,9	12,9	0	0		STŘEDNÍ RIZIKO	6.12.2005															
20	17	3	3	0	0	5	5	1	1	0	0		VYSOKÉ RIZIKO	20.1.2006			X			X									
21	18	2a	2	0	0	7	7	157	157	0	0		VYSOKÉ RIZIKO	12.12.2005			X			X									
22	19	3b	3	0	0	4	4	5,8	5,8	0	0		VYSOKÉ RIZIKO	15.12.2005										1					
23	20	1c	1	0	0	5	5	14	14	0	0		STŘEDNÍ RIZIKO	17.1.2006			X			X									
24	21	1c	1	0	0	6	6	8,9	8,9	0	0		STŘEDNÍ RIZIKO	21.12.2005								C19	30.1.2007	1					
25	22	1c	1	0	0	6	6	8,9	8,9	0	0		NÍZKÉ RIZIKO	13.1.2006										2					

Obrázek 4: Klasifikace pacientů do tříd

Zbývá zmínit ještě jednu poznámku. Nezapomínejme na **protokol studie**. Klinické studie trvají mnohdy několik let. Je známo, že ve farmaceutickém průmyslu dochází k vysoké fluktuaci pracovníků. Snadno se může stát, že se pracovní tým na obou stranách projektu v průběhu studie několikrát významně obmění. V takovémto případě velmi oceníme dohledatelnost všech výsledků a všech postupů. Protokol studie by měl obsahovat informaci o zadávání dat, o jejich validaci, popis randomizace (dvojitě zaslepení), definice klíčových parametrů, definice populace či populací, na nichž je analýza prováděna a nezapomeňme ani na plánování interim (definovaných průběžných) analýz (minimální počet pacientů, minimální doba sledování, apod.), návrh rozsahu výběru, atd.

## 4.Exploratorní analýza proměnných

Původním posláním statistiky bylo zjišťování údajů o populaci na základě výběrového souboru. Pod pojmem **populace** přitom rozumějte souhrn všech existujících prvků, které sledujeme při statistickém výzkumu.

Vzhledem k tomu, že rozsah (počet prvků) populace je obvykle vysoký, provádí se většinou tzv. **výběrová šetření**, kdy se namísto celé populace zkoumá pouze její část. Zkoumaná část populace se nazývá **výběr**, popř. výběrový soubor. Otázkou je jak stanovit takový výběr, aby byl skutečně reprezentativní, tj. aby parametry výběru (např. průměr) dostatečně přesně reprezentovaly parametry populace.

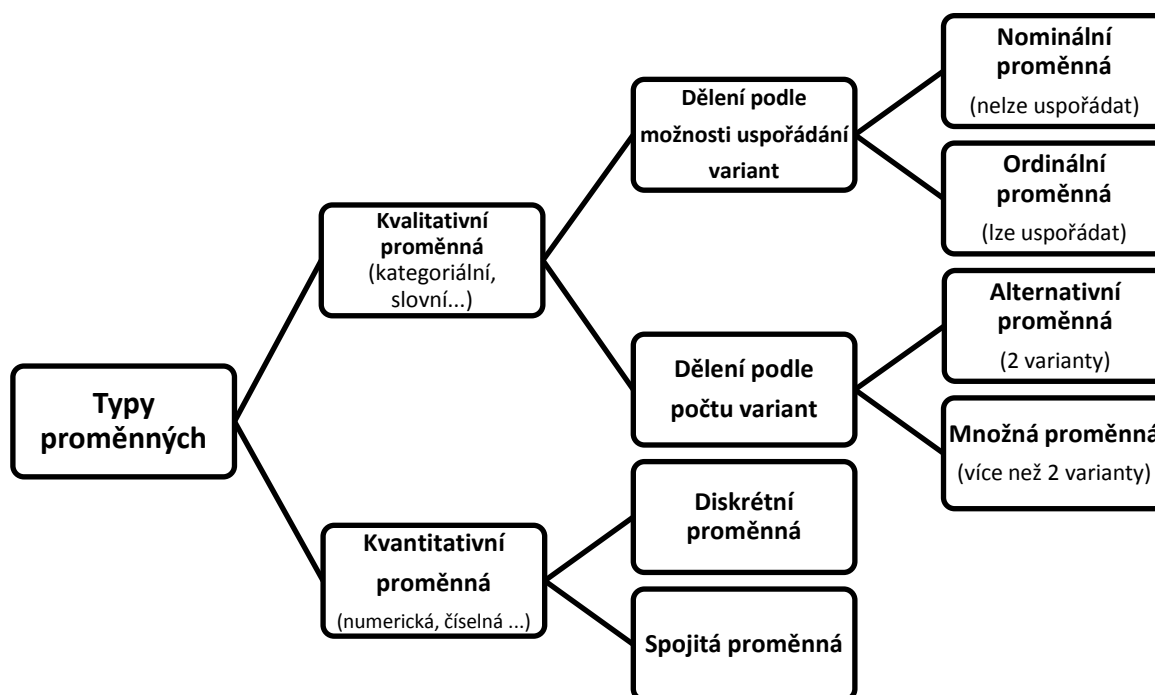
Existuje několik způsobů jak výběr provést. Abychom se vyvarovali opomenutí některých prvků populace, zvolíme tzv. **náhodný výběr**, v němž každý prvek populace má stejnou šanci být zařazen do výběru.

Je zřejmé, že výběrové šetření nemůže být nikdy tak přesné jako průzkum celé populace. Proč jej tedy preferujeme?

1. Úspora času a finančních prostředků (zejména u rozsáhlé populace)
2. Destruktivní testování (některé testy – pevnost lan, životnost zářivek, obsah cholesterolu v krvi, atd. – vedou k destrukci zkoumaných prvků; zamyslete se sami, k čemu by vedlo testování celé populace)
3. Nedostupnost celé populace (při srovnávání působení faktorů okolí a dědičných znaků poskytují nejlepší informace identická dvojčata – jak je všechna sehnat a přesvědčit ke spolupráci?)

Statistikové dokážou popsat celou populaci na základě poznatků z výběru, proto přejdeme k základním výběrovým šetřením neboli k exploratorní analýze (exploratory data analysis – EDA). Údajům, které u souboru sledujeme, budeme říkat **proměnné** a jejich jednotlivým hodnotám **varianty proměnné**. **Exploratorní (popisná) statistika** bývá prvním krokem k odhalení informací skrytých ve velkém množství proměnných a jejich variant. To znamená uspořádání proměnných do názornější formy a jejich popis několika málo hodnotami, které by obsahovaly co největší množství informací obsažených v původním souboru.

Vzhledem k tomu, že způsob zpracování proměnných závisí především na jejich typu, seznámíme se nyní se základním dělením proměnných do různých kategorií. Toto dělení je prezentováno na následujícím obrázku:



Obrázek 5: Schéma typu proměnných v exploratorní analýze

- **Proměnná kvalitativní** – její varianty jsou vyjádřeny slovně a podle vztahu mezi jednotlivými hodnotami se dělí na dvě základní podskupiny:
  - **Proměnná nominální (jmenná)** – nabývá rovnocenných variant; nelze je porovnávat ani seřadit (např. pohlaví, národnost, značka hodinek...)
  - **Proměnná ordinální** – tvoří přechod mezi kvalitativními a kvantitativními proměnnými; jednotlivým variantám lze přiřadit pořadí a vzájemně je porovnávat nebo seřadit (např. známka ve škole, velikost oděvů (S, M, L, XL))

Jiným způsobem dělení kvalitativních proměnných je dělení podle počtu variant, jichž proměnné mohou nabývat:

- **Proměnná alternativní** – nabývá pouze dvou různých variant (např. pohlaví...)
  - **Proměnná množná** – nabývá více než dvou různých variant (např. vzdělání, jméno, barva očí...)
- **Proměnná kvantitativní** – je vyjádřena číselně a dělí se na:
  - **Proměnná diskrétní** – nabývá konečného nebo spočetného množství variant (např. známka z matematiky)
    - **Proměnná diskrétní konečná** – nabývá konečného počtu variant (např. známka z matematiky)
    - **Proměnná diskrétní spočetná** – nabývá spočetného množství variant (např. věk v letech, výška v centimetrech, váha v kilogramech...)
  - **Proměnná spojitá** – nabývá libovolné hodnoty z  $\mathbb{R}$  nebo z nějaké podmnožiny  $\mathbb{R}$  (např. výška, hmotnost, vzdálenost měst...)

## 5. Kontingenční tabulka a její testy

Testy nezávislosti v kontingenční tabulce (Contingency Tables, Crosstables) řadíme mezi tzv. analýzu kategoriálních dat (Categorical Data Analysis). Setkáváme se s nimi v ekonomii, personalistice, psychologii, sociologii, marketingu...

**Kontingenční tabulka** vzniká setříděním prvků populace podle variant dvou kategoriálních znaků, např. znaku X a znaku Y. Necht' znak X má  $m$  variant a znak Y má  $n$  variant. Názvy jednotlivých variant znaků X a Y jsou pak uvedeny v hlavičce tabulky a uvnitř tabulky uvádíme četnosti  $n_{ij}$ , kde  $i$  označuje  $i$ -tou variantu znaku X ( $i \in \langle 1; m \rangle$ ) a  $j$  označuje  $j$ -tou variantu znaku Y ( $j \in \langle 1; n \rangle$ ). Při práci s kontingenční tabulkou budeme dále používat toto značení:

$n_i$  ... součet všech četností v  $i$ -té řádce

$n_j$  ... součet všech četností v  $j$ -tém sloupci

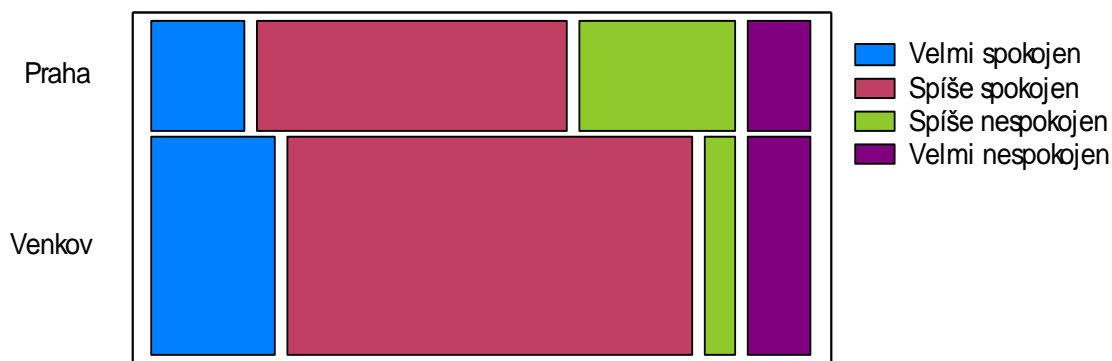
### Schéma kontingenční tabulky

X/Y	$Y_1$	$Y_2$	...	$Y_n$	$\Sigma_j$
$X_1$	$n_{11}$	$n_{12}$	$\vdots$	$n_{1n}$	$n_{1.}$
$X_2$	$n_{21}$	$n_{22}$	$\vdots$	$n_{2n}$	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$X_m$	$n_{m1}$	$n_{m2}$	...	$n_{mn}$	$n_{m.}$
$\Sigma_i$	$n_{.1}$	$n_{.2}$	...	$n_{.n}$	$n$

Tabulka 1: Schéma kontingenční tabulky

Grafickou obdobou kontingenční tabulky je **mozaikový graf**. Tento graf se skládá z obdélníků, jejichž strany jsou úměrné příslušným marginálním relativním četnostem. Statgraphics Plus konstruuje mozaikový graf tak, že na svislou osu vynáší nezávisle proměnnou (příčina) a na vodorovnou osu závisle proměnnou (důsledek). Pokud by byl mozaikový graf v tomto případě tvořen svislými pruhy (jednotlivé obdélníky stejných barev by měly stejné „vodorovné“ rozměry), znamenalo by to, že sledované proměnné jsou nezávislé. Obdobné vyhodnocení provedeme v případě, kdy statistický software vynáší nezávisle proměnnou na vodorovnou osu (např. JMP-IN). Pak je v případě nezávislosti proměnných mozaikový graf tvořen vodorovnými pásy.

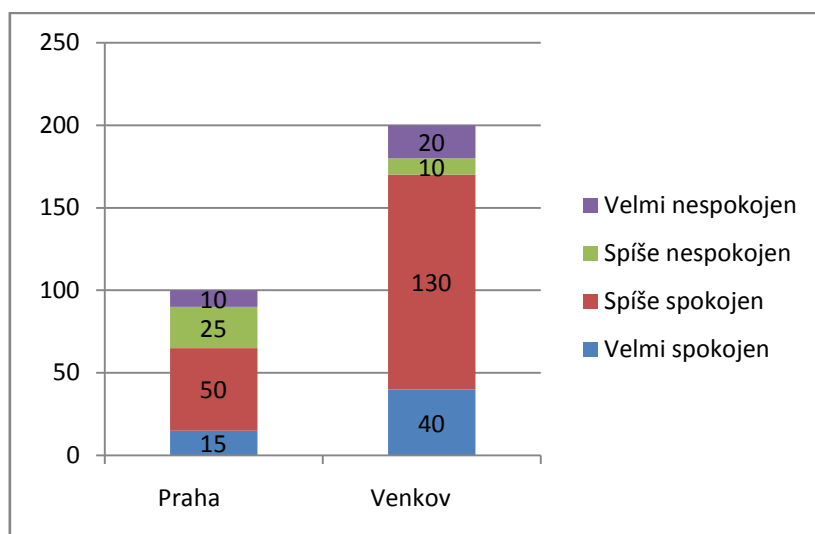
Následující ukázka mozaikového grafu odpovídá datům popisujícím, jak jsou zaměstnanci spokojeni v práci v závislosti na umístění podniku.



Obrázek 6: Mozaikový graf

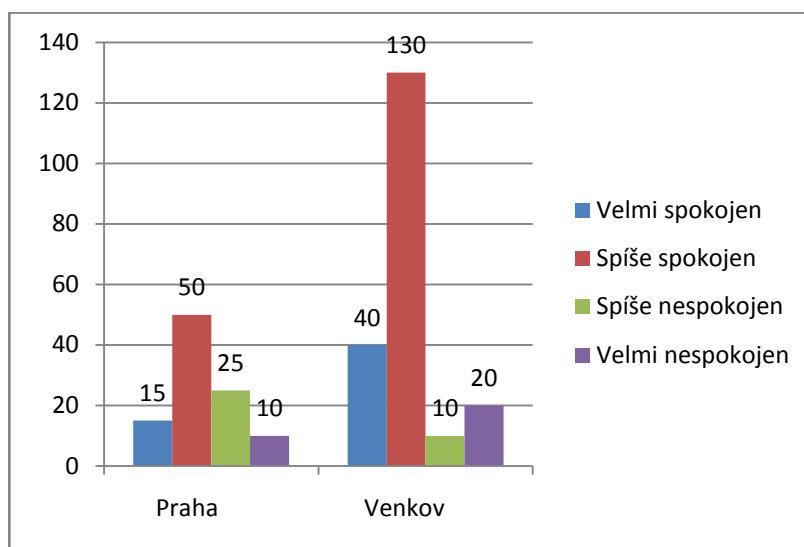
Obdobou mozaikového grafu je **100% skládaný pruhový graf** (např. MS Excel). Od mozaikového grafu se tento graf liší tím, že šířky obou řádků jsou stejné, tzn., že nezohledňuje řádkové marginální relativní četnosti.

Kromě mozaikového grafu se pro prezentaci dat zapsaných v kontingenční tabulce používají **shlukový**, popř. **kumulativní sloupcový graf**.



Obrázek 7: Kumulativní sloupcový graf





Obrázek 8: Shlukový graf

## Testy v kontingenční tabulce

Pro ověření nezávislosti náhodných veličin  $X$  a  $Y$  (nezávislosti v kombinační tabulce) používáme test, který je založen na **porovnávání empirických** (pozorovaných) **četností** s **četnostmi teoretickými**, tj. takovými, které bychom očekávali v případě nezávislosti.

Teoretické četnosti označujeme  $n_{ij}^*$  a určujeme je jako četnosti odpovídající součinu příslušných marginálních relativních četností (připomeňme si, že v případě, že jsou dvě diskrétní náhodné veličiny nezávislé, pak jejich sdružené pravděpodobnosti jsou rovny součinu příslušných marginálních pravděpodobností).

$$n_{ij}^* = \left( \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} \right) \cdot n = \frac{n_{i.} \cdot n_{.j}}{n}$$

## $\chi^2$ test nezávislosti v kontingenční tabulce

### 1. Volba nulové a alternativní hypotézy

**H<sub>0</sub>**: Náhodné veličiny v kombinační tabulce jsou nezávislé.

**H<sub>A</sub>**: Náhodné veličiny v kombinační tabulce jsou závislé.

### 2. Volba testové statistiky $T(\underline{X})$ (včetně nulového rozdělení)

$$T(\underline{X}) = G = \sum_{i=1}^m \sum_{j=1}^n \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} \rightarrow \chi^2_{(m-1)(n-1)}$$

Testová statistika  $G$  má rozdělení  $\chi^2$  s  $(m-1) \cdot (n-1)$  stupni volnosti. Je zřejmé, že čím bude hodnota testové statistiky  $G$  vzdálenější od nuly, tím silněji budou data vypovídat pro zamítnutí nulové hypotézy.

**Předpoklad testu:**

1. Žádná z očekávaných četností nesmí klesnout pod hodnotu 2.
2. Alespoň 80% očekávaných četností musí být větších než 5.

**3. Výpočet p-value**

Při tomto testu určujeme p-value jako:  $p - value = 1 - F_0(\chi_{OBS})$

## 6. Cenzorovaná data

### Výběrové plány

Předpokládejme, že v okamžiku  $t = 0$  začneme za účelem zjištění charakteristik spolehlivosti pozorovat  $n$  pacientů stejného typu onemocnění a že jejich doba přežití má (stejnou) diskrétní funkci  $F$ . Klasická statistická situace nastává, jestliže pozorování provádíme tak dlouho, dokud všichni pacienti nezemřou nebo nenastane nějaká další situace (přerušení léčby, odstěhování pacienta, nastane remise apod.). Výsledkem experimentu je potom úplný výběr  $X_1, \dots, X_n$  dob přežití, tj. náhodný výběr v běžném slova smyslu.

V aplikacích jako např. při zkouškách životnosti složitých systémů, při klinických pokusech, ale i v pojišťovnictví je experiment často analyzován předtím, než se všechny prvky porouchají. Tak je tomu v teorii spolehlivosti, když některé prvky jsou vyjmuty ze sledování před ukončením experimentu, a v klinických pokusech dochází k úmrtím z příčin, které nejsou předmětem zkoumání nebo se pacienti prostě odstěhují.

Ve všech takových a podobných případech má statistika k dispozici **neúplné výběry**. S některými základními uspořádáními experimentu vedoucími k neúplným výběrům se nyní seznámíme. Opět budeme předpokládat, že pozorujeme  $n$  pacientů se stejným onemocněním; jejich doby přežití označíme  $X_1, \dots, X_n$ . Ne všechna  $X_i$  však jsou opravdu pozorovatelná!

### Cenzorování typu I (cenzorování časem)

Předepíšeme číslo  $T > 0$  zvané časový cenzor a v okamžiku  $t = 0$  zahájíme pozorování. Experiment ukončíme v okamžiku  $T$  bez ohledu na to, kolik pacientů skutečně zemřelo. Necht'  $X_{(1)} \leq \dots \leq X_{(n)}$  označujeme uspořádaný náhodný výběr z  $X_1, \dots, X_n$ . Výsledkem experimentu je potom prvních řekněme  $r$  hodnot pořádkových statistik  $X_{(1)} \leq \dots \leq X_{(r)} \leq T$  a informace  $X_{(r+1)} > T$ . Počet skutečně pozorovaných úmrtí  $r$  je zde náhodná veličina, která může nabývat hodnot  $0, 1, \dots, n$ . Doba trvání experimentu je pevná a rovna  $T$ .

### Cenzorování typu II (cenzorování poruchou)

Předepíšeme přirozené číslo  $r \leq n$  a v okamžiku  $t = 0$  zahájíme pozorování. Experiment ukončíme v okamžiku úmrtí  $r$ -tého pacienta. Výsledkem experimentu je potom prvních  $r$  hodnot pořádkových statistik  $X_{(1)} \leq \dots \leq X_{(r)}$ . Počet skutečně pozorovaných úmrtí je předem určen experimentátorem a roven  $r$ . Doba trvání experimentu je náhodná veličina daná okamžikem úmrtí  $r$ -tého pacienta, tedy  $X_{(r)}$ .

### Náhodné cenzorování

Při sledování složitých systémů nemáme většinou možnost uspořádat experiment ideálně podle našich představ a musíme využít kupř. provozních dat, tj. pozorování ze skutečného provozu. Zjistíme, že se často časové cenzory liší systém od systému, a že ukončení pozorování u každého systému je více méně náhodné. Skončí-li pozorování systému  $1, \dots, n$  v okamžicích  $T_1, \dots, T_n$ , je ekonomicky neúnosné uvažovat je jako cenzorovaná v okamžiku  $T = \min(T_1, \dots, T_n)$ . Výhodnější je vybudovat aparát pro případ, kdy pozorování každého systému je ukončeno v náhodném okamžiku  $T$ . V takovém případě mluvíme o náhodném cenzorování. Spolu s náhodnou veličinou  $X$  reprezentující

dobu přežití uvažujeme náhodnou veličinu  $T$  reprezentující časový cenzor. U každého pacienta pak pozorujeme buď  $X$  nebo  $T$  podle toho, zda dříve nastalo úmrtí nebo byl pacient dříve vyjmut ze sledování. Navíc máme k dispozici informaci, zda nastalo úmrtí nebo zda byl pacient vyjmut ze sledování. Výsledkem experimentu je tedy  $n$  dvojic

$$(W_1, I_1), \dots, (W_n, I_n),$$

kde

$$W_j = \min(X_j, T_j),$$

$I_j = 1$ , jestliže  $W_j = X_j$  ( $j$ -té pozorování je **necenzorované**)

$I_j = 0$ , jestliže  $W_j = T_j$  ( $j$ -té pozorování je **cenzorováno** v čase  $T_j$ ),

$$j = 1, \dots, n.$$

Jedná se tedy o náhodný výběr z dvourozměrného rozdělení.

## 7. Neparametrické metody

Přestože parametrické metody zaujímají klíčovou úlohu ve statistické analýze spolehlivosti, je možné některé problémy řešit i při neparametrickém přístupu. Budou zde uvedeny neparametrické odhady funkce spolehlivosti, pomocí kterých lze odhadnout momenty doby přežití a některé jiné charakteristiky. Nevýhodou všech neparametrických odhadů je však nemožnost získat informaci o chování chvostů rozdělení. Na druhou stranu výhodou těchto odhadů, je možnost jejich sestavení z poměrně „divokých“ dat. Tak můžeme vyřešit nejednu situaci, se kterou se v praxi setkáme.

Uvedené odhady se často používají nejen v teorii spolehlivosti, ale i v klinickém výzkumu a pojišťovnictví při analýze tabulek úmrtnosti.

Nechť  $X$  označuje dobu přežití,  $R(x)$  odpovídající funkci spolehlivosti,  $0 = x_0 < x_1 < \dots < x_k$  „význačné“ časové okamžiky a  $J_i = (x_{i-1}, x_i)$ ,  $i = 1, \dots, k$ . Předpokládejme, že experiment je uspořádán tak, že v okamžiku  $x_0 = 0$  začneme pozorovat  $n$  pacientů se stejným onemocněním, a že údaje o průběhu experimentu můžeme zjišťovat pouze v časových okamžicích  $x_1, \dots, x_k$ . Získané údaje jsou dejme tomu

$n_i$  ... počet žijících pacientů a sledovaných do okamžiku  $x_{i-1}$  (okamžik  $x_{i-1}$  v to nepočítaje),

$d_i$  ... počet úmrtí v intervalu  $J_i$ ,

$v_i$  ... počet pacientů, kteří se „ztratili“ ze sledování (odstěhovali se) v intervalu  $J_i$

$w_i$  ... počet pacientů, kteří byli záměrně vyjmuti ze sledování,  $i = 1, \dots, k$ .

Položme  $n_0 = n$ ,  $d_0 = v_0 = w_0 = 0$ .

Pro  $t \in J_i$  můžeme hodnotu  $R(t)$  odhadnout pomocí

$$\text{est } R(t) = 1 - d_{i-1} / n_{i-1}.$$

Tento odhad (někdy nazývaný odhadem založeným na redukovaném výběru) však ignoruje informaci obsaženou ve  $v_i$  a  $w_i$ . Myšlenka zahrnout i tuto informaci je založena na následující pravděpodobnostní úvaze. Označme  $E_i = \{x \geq x_i\}$ ,  $p_i = P(E_i | E_{i-1})$ ,  $i = 1, \dots, k$ . Potom zřejmě platí

$$(7.1) \quad P(E_k) = P(E_k | E_{k-1})P(E_{k-1}) = \dots = P(E_k | E_{k-1})P(E_{k-1} | E_{k-2}) \dots P(E_1) = \prod_{i=1}^k p_i.$$

V případě, že v intervalu  $J_i$  nedošlo ke „ztrátám“, a že žádní pacienti nebyli vyjmuti ze sledování, můžeme  $p_i$  odhadnout pomocí veličiny  $1 - d_i/n_i$ . V případě, že v intervalu  $J_i$  došlo ke „ztrátám“ nebo záměrnému vyjmutí ze sledování, se předpokládá, že „ztracení“ a vyjmutí pacienti byli sledování polovinu příslušného intervalu. Potom počítáme s tzv. **efektivním počtem pacientů sledovaných v intervalu  $J_i$** .

$$(7.2) \quad n_i = n_i - \frac{1}{2}(v_i + w_i).$$

Za odhad podmíněných pravděpodobností  $p_i$  potom vezmeme

$$(7.3) \quad \tilde{p}_i = 1 - d_i/n_i,$$

což spolu se (7.1) vede k odhadu funkce spolehlivosti

$$(7.4) \quad \tilde{R}(t) = \prod_{i=1}^j \tilde{p}_i, \quad x_j < t \leq x_{j+1}, \quad j=1, \dots, k-1,$$

$$\tilde{R}(t) = 1, \quad t \leq x_1.$$

Tabulky úmrtnosti jsou obvykle konstruovány s ekvidistantními časovými intervaly, přičemž délka časového intervalu je jeden rok. Abychom dostali data se strukturou výše uvedenou, museli bychom sledovat určitou skupinu lidí (v této souvislosti zvanou **kohorta**) po několik desetiletí. Většinou se však používají průběžné tabulky úmrtnosti, které vzniknou ze statistik během jednoho roku, a kdy skupiny jedinců v disjunktních časových intervalech jsou disjunktní.

## Kaplan-Meierův odhad (angl. Product limit estimate)

Je to v současné době nejpoužívanější neparametrický odhad funkce spolehlivosti. Vychází opět z vyjádření (7.1), ale za „význačné“ okamžiky bere přímo okamžiky, kdy pacient zemřel nebo byl vyjmut ze sledování. Podobně jako (7.4), ani Kaplan – Meierův odhad nerozlišuje pacienty, kteří se „ztratili“, a pacienty, kteří byli vyjmuti ze sledování. Můžeme proto předpokládat, že data jsou náhodně cenzorována a výsledkem experimentu je  $n$  dvojic

$$(W_1, I_1), \dots, (W_n, I_n),$$

kde  $W_j$  je okamžik poruchy resp. vyjmutí  $j$ -tého prvku ze sledování a  $I_j = 1$  resp.  $I_j = 0$  podle toho, zda dříve došlo k úmrtí resp. vyjmutí. Předpokládejme pro jednoduchost, že ve výběru  $W_1, \dots, W_n$  nedošlo ke shodám, a utvořme uspořádaný náhodný výběr  $W_{(1)} < \dots < W_{(n)}$ . Necht'  $I_{(j)}$  je indikátor odpovídající  $W_{(j)}$ ,  $j = 1, \dots, n$ . (Pozor,  $I_{(1)}, \dots, I_{(n)}$  **nej**sou uspořádána!). Za význačné okamžiky vezmeme  $W_{(1)}, \dots, W_{(n)}$ .

Označme nyní

$n_i$  ... počet přežitých pacientů do okamžiku  $W_{(i)}$  (okamžik  $W_{(i)}$  v to nepočítaje),

$d_i$  ... počet úmrtí v okamžiku  $W_{(i)}$ .

Potom za odhady podmíněných pravděpodobností  $p_i$  vezmeme

$$(7.5) \quad \tilde{p}_i = 1 - d_i/n_i, \quad i = 1, \dots, n.$$

Poznamenejme, že  $\tilde{p}_i = 1 - 1/n_i$ , jestliže  $I_{(i)} = 1$  a  $\tilde{p}_i = 1$ , jestliže  $I_{(i)} = 0$ .

**Kaplan-Meierův odhad** funkce spolehlivosti je potom

$$(7.6) \quad \tilde{R}(t) = \prod_{i: W_{(i)} < t} \tilde{p}_i, \quad t \leq W_{(n)},$$

$$\tilde{R}(t) = 0, \quad t > W_{(n)}.$$

(Prázdný součin definujeme jako rovný jedné.) Alternativní tvar Kaplan-Meierova odhadu je

$$(7.7) \quad \tilde{R}(t) = \prod_{i: W_{(i)} < t} \left(1 - \frac{1}{n_i}\right)^{I_{(i)}} = \prod_{i: W_{(i)} < t} \left(\frac{n-i}{n-i+1}\right)^{I_{(i)}}, \quad t \leq W_{(n)},$$

$$\tilde{R}(t) = 0, \quad t > W_{(n)}.$$

Reálná data mohou obsahovat shody. V takovém případě modifikujeme Kaplan-Meierův odhad následujícím způsobem. Nechť  $R_i$  označuje pořadí dvojic  $(W_i, 1 - I_i)$  v lexikografickém uspořádání posloupnosti

$$(W_1, 1 - I_1), \dots, (W_n, 1 - I_n).$$

Potom modifikovaný Kaplan-Meierův odhad je

$$(7.8) \quad \tilde{R}(t) = \prod_{i: W_{(i)} < t} \left( \frac{n - R_i}{n - R_i + 1} \right)^{I_i}, t \leq W_{(n)},$$

$$\tilde{R}(t) = 0, \quad t > W_{(n)}.$$

Poznamenejme, že pro úplný výběr je Kaplan-Meierův odhad totožný s empirickou funkcí spolehlivosti.

Asymptotické vlastnosti Kaplan-Meierova odhadu v případě náhodného cenzorování jsou uvedeny v následující větě.

**Věta 1:** Nechť distribuční funkce  $F$  doby přežití  $X$  a distribuční funkce  $G$  časového cenzoru jsou spojité. Nechť  $t > 0$  je takové, že  $R(t) = 1 - F(t) > 0$ . Potom

$$(7.9) \quad \sqrt{n} (\hat{R}(t) - R(t)) \xrightarrow{D} N(0, R^2(t) \int_0^t ((1 - F(x))(1 - G(x)))^{-2} dP(X < x, I = 1)).$$

**Bez důkazu.**

V praxi je třeba nahradit rozptyl asymptotického rozložení ve větě 7.1 nějakým odhadem. Ukážeme jeden z možných postupů. Předně je patrné, že pravděpodobnost  $P(X < x, I = 1)$  je možné odhadnout pomocí relativní četnosti jako

$$\hat{P}(X < x, I = 1) = \frac{1}{n} \sum_{i: W_i < x} I_i,$$

takže  $\hat{P}$  jakožto funkce  $x$  má skoky velikosti  $1/n$  v bodech  $W_i$  s  $I_i = 1$ . Dále  $H(x) = 1 - (1 - F(x))(1 - G(x))$  je distribuční funkce náhodné veličiny  $W$ , takže  $H$  můžeme odhadnout pomocí obyčejné empirické distribuční funkce založené na výběru  $W_1, \dots, W_n$ :

$$\hat{H}(x) = \frac{1}{n} \sum_{i=1}^n I(W_i < x).$$

Vzhledem k chování  $\hat{P}$  potřebujeme znát odhady  $\hat{H}$  pouze v bodech  $W_{(i)}$ . Místo  $(1 - \hat{H}(x))^2$  v (7.9) použijeme „symetrizovaný“ odhad

$$(1 - \hat{H}(x))(1 - \hat{H}(x+)).$$

Vzhledem k tomu, že  $\hat{H}(W_i) = \frac{i-1}{n}$ ,  $H(W_{(i)} +) = \frac{i}{n}$ , je možné odhadnout rozptyl v (7.9) pomocí

$$\hat{R}^2(t) \sum_{i: W_{(i)} < t} ((1 - \hat{H}(W_{(i)}))(1 - \hat{H}(W_{(i)} +)))^{-2} \frac{I_{(i)}}{n},$$

z čehož

$$(7.10) \quad \widehat{\text{var}} \hat{R}(t) = \hat{R}^2(t) \sum_{i: W_{(i)} < t} \frac{I_{(i)}}{(n-i)(n-i+1)} - \text{vzorec je znám jako GREENWOODova formule}$$

## 8. Vlastní zpracování dat

V této kapitole se budu zabývat vlastním řešením jednotlivých otázek, které jsem zpracovávala pro onkologické oddělení Fakultní nemocnice Ostrava.

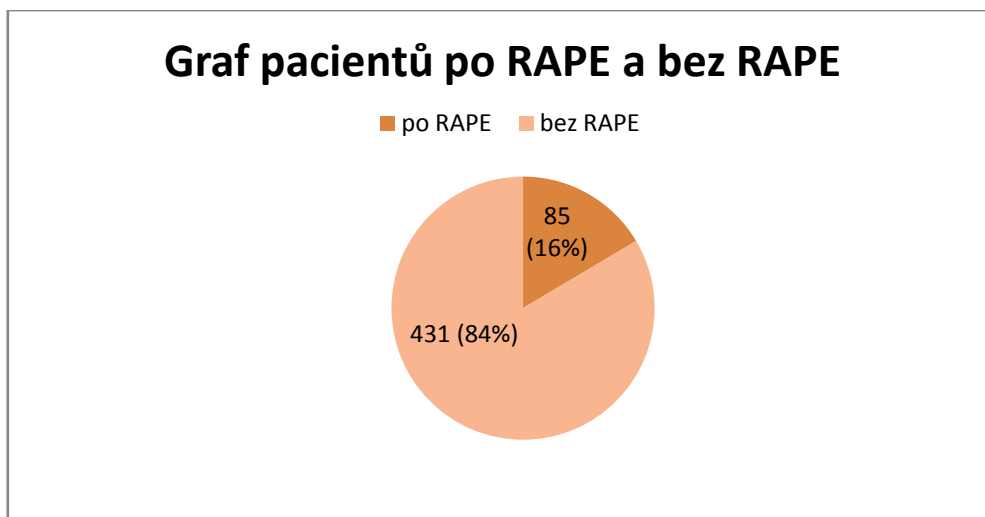
Pro zpracování výsledků byly použity data 502 pacientů, přičemž v jednotlivých bodech byli použiti pouze ti pacienti, kteří měli uvedeny potřebné údaje ke zpracování.

U některých výsledků jsem zároveň zjišťovala pomocí kontingenčních tabulek závislosti mezi jednotlivými proměnnými. Zde uvádím pouze zlomek ze zpracovaných grafů. Celkové zpracování je uvedeno v příloze [II.].

### Jednotlivé řešení otázky

1. Kolik pacientů v jednotlivých stádiích podstoupilo radioterapii –K resp.
2. Kolik pacientů v jednotlivých stádiích podstoupilo RAPE či LAE či kombinaci, OE
3. Kolik pacientů po RAPE mělo radioterapii K-A, kolik K-S.
4. Kolik pacientů v jednotlivých stádiích mělo nějaký typ hormonální léčby
5. Jaké je procentuální zastoupení jednotlivých stupňů akutní a pozdní toxicity (akutní RTOG a pozdní RTOG) po RAPE a bez RAPE
6. Zpracovat křivky přežití pro skupiny M0 a M1

Nejprve základní graf ukazující množství pacientů, kteří podstoupili RAPE a kteří ne.



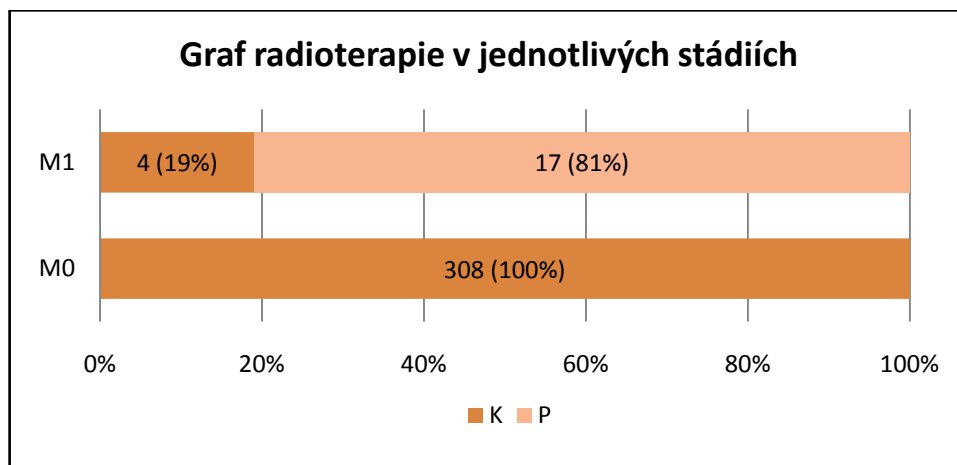
Obrázek 9: Graf pacientů po a bez RAPE

Operace	Množství pacientů
po RAPE	85
bez RAPE	431



## 1. Kolik pacientů v jednotlivých stádiích podstoupilo radioterapii –K resp.

Ve skupině M0 se nevyskytují žádní pacienti, kteří podstoupili **paliativní** radioterapii. Za to ve skupině M1 z 21 pacientů v téhle skupině podstoupili pouze 4 radioterapii **kurativní**.



Obrázek 10: Graf radioterapie v jednotlivých stádiích

Skupina	M0	M1
K	308 (100%)	4 (19%)
P	0	17 (81%)

Určení závislosti proměnných pomocí kontingenční tabulky:

Kontingenční tabulka závislosti rizikové skupiny na typu radioterapie:

Rizik. Skup.	K	P	Celkový součet
M0	308		308
M1	4	17	21
Celkový součet	312	17	329

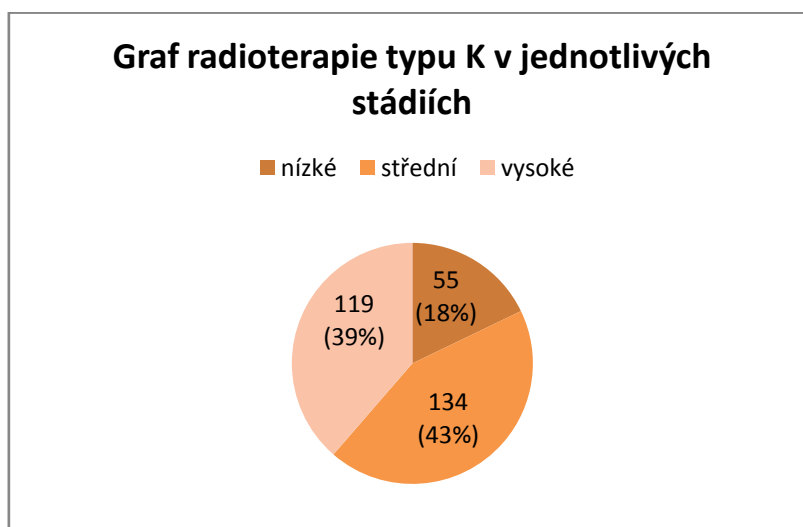
Tabulka 2: Závislost rizikové skupiny na typu radioterapie

Volba nulové a alternativní hypotézy:

**H0:** Riziková skupina pacientů nezávisí na typu radioterapie.

**HA:** Riziková skupina pacientů závisí na typu radioterapie.

p-value =  $3,97 \cdot 10^{-59}$ , proto zamítáme nulovou hypotézu ve prospěch alternativní hypotézy → **NELZE TVRDIT, že riziková skupina pacientů nezávisí na typu radioterapie.**



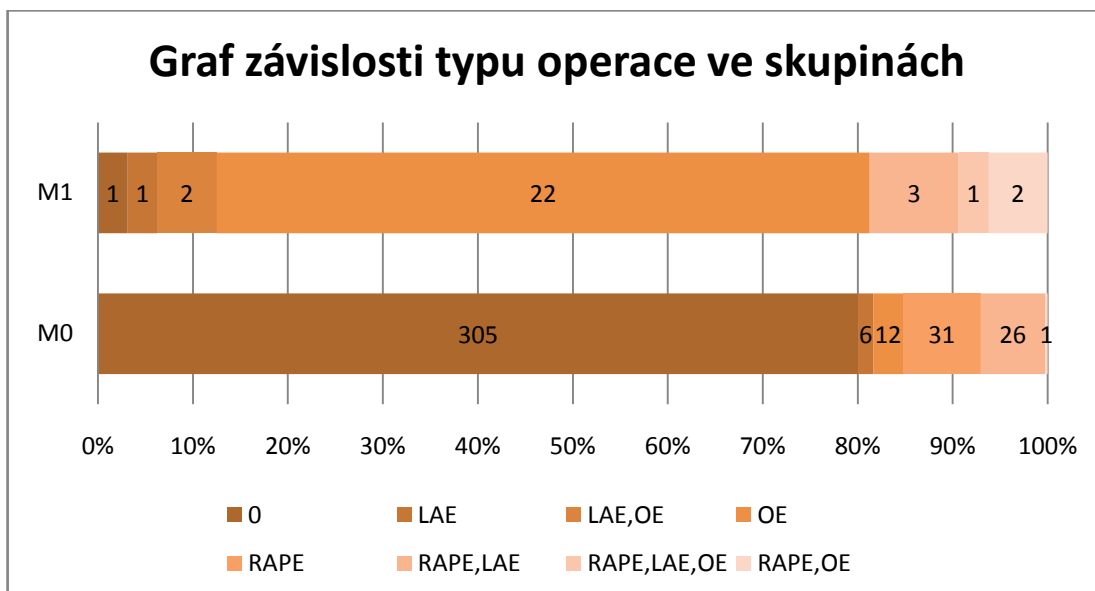
Obrázek 11: Graf radioterapie typu K v jednotlivých stádiích

typ radioterapie	nízké	střední	vysoké
<b>K</b>	55 (18%)	134 (43%)	119 (39%)

**Kurativní** radioterapii podstoupila většina pacientů ve středním a vysokém stádiu a nejméně pacientů ve stádiu nízkém.

## 2. Kolik pacientů v jednotlivých stádiích podstoupilo RAPE či LAE či kombinaci, OE

Ve skupině M0 se vyskytuje více než 80% pacientů bez operace zatímco ve skupině M1 jsou nejvíce zastoupeni pacienti s operací OE.



Obrázek 12: Graf závislosti typu operace ve skupinách

Typ operace	M0	M1
0	305 (80,1%)	1 (3,1%)
LAE	6 (1,6%)	1 (3,1%)
LAE,OE	0	2 (6,3%)
OE	12 (3,1%)	22 (68,7%)
RAPE	31 (8,1%)	0
RAPE,LAE	26 (6,8%)	3 (9,4%)
RAPE,LAE,OE	0	1 (3,1%)
RAPE,OE	1 (0,3%)	2 (6,3%)

Určení závislosti proměnných pomocí kontingenční tabulky:

Kontingenční tabulka závislosti rizikové skupiny na typu operace:

Rizik. Skup.	0	LAE+LAE,OE+OE	RAPE+jeho kombinace	Celkový součet
M0	305	18	58	381
M1	1	25	6	32
Celkový součet	306	43	64	413

Tabulka 3: Závislost rizikové skupiny na typu operace

Volba nulové a alternativní hypotézy:

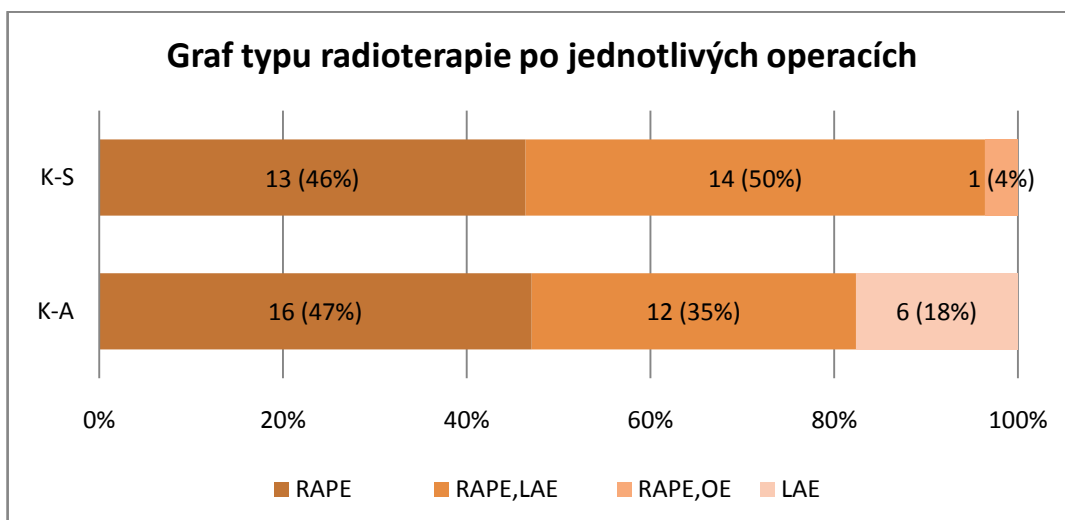
**H0:** Riziková skupina pacientů nezávisí na typu operace.

**HA:** Riziková skupina pacientů závisí na typu operace.

p-value =  $4,54 \cdot 10^{-39}$ , proto zamítáme nulovou hypotézu ve prospěch alternativní hypotézy → **NELZE TVRDIT, že riziková skupina pacientů nezávisí na typu operace.**

### 3. Kolik pacientů po RAPE mělo radioterapii K-A, kolik K-S.

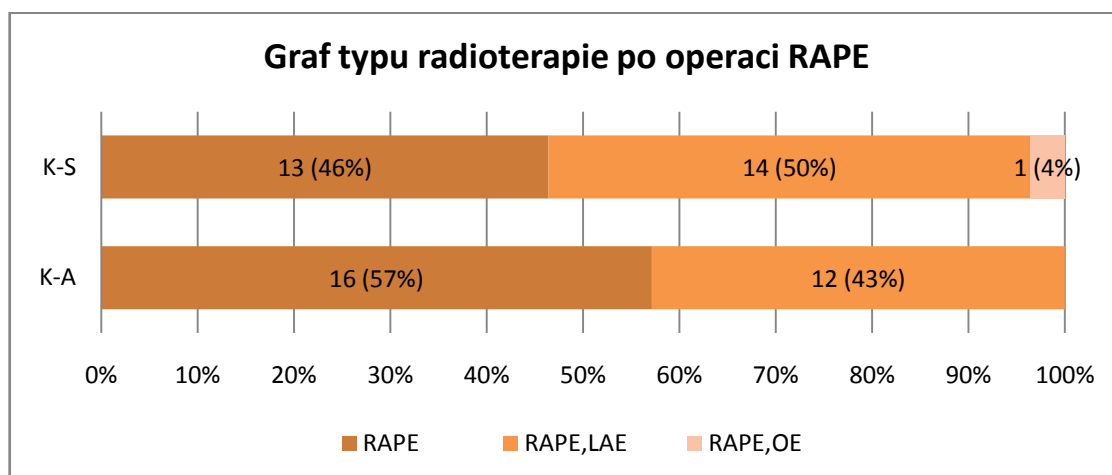
Na grafu vidíme, jak vypadá rozložení pacientů, kteří podstoupili radioterapii K-A nebo K-S po jednotlivých operacích. Je zde vidět, že i když se radioterapie typu K-A a K-S týkají hlavně pacientů po RAPE, vyskytují se zde i pár pacientů, kteří podstoupili K-A po operaci LAE.



Obrázek 13: Graf typu radioterapie po jednotlivých operacích

typ operace	K-A	K-S
RAPE	16 (47%)	13 (46%)
RAPE,LAE	12 (35%)	14 (50%)
RAPE,OE	0	1 (4%)
LAE	6 (18%)	0

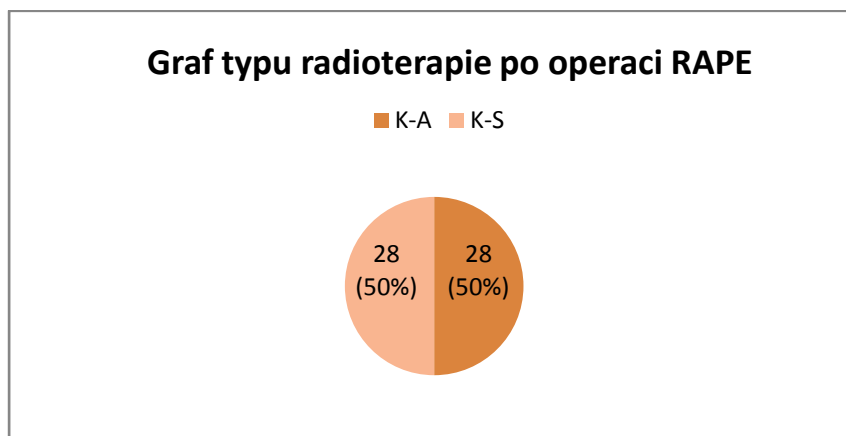
Následující graf již ukazuje pouze pacienty, kteří podstoupili RAPE.



Obrázek 14: Graf typu radioterapie po operaci RAPE

typ operace	K-A	K-S
<b>RAPE</b>	16 (57%)	13 (46%)
<b>RAPE,LAE</b>	12 (43%)	14 (50%)
<b>RAPE,OE</b>	0	1 (4%)

Co se týče četnosti K-A a K-S po operaci RAPE, je to 50% na 50%. V obou typech radioterapie jsou rovnoměrně rozdělení pacienti, kteří podstoupili operaci RAPE a RAPE s kombinací LAE.



Obrázek 15: Výšečový graf typu radioterapie po operaci RAPE

**Určení závislosti proměnných pomocí kontingenční tabulky:**

Kontingenční tabulka závislosti typu radioterapie na typu operace:

Radioterapie	0	LAE+LAE,OE+OE	RAPE+jeho kombinace	Celkový součet
<b>K</b>	321	19		340
<b>K-A</b>		7	34	41
<b>K-S</b>			44	44
<b>P</b>	1	44	3	48
<b>Celkový součet</b>	322	70	81	473

Tabulka 4: Závislost typu radioterapie na typu operace

Volba nulové a alternativní hypotézy:

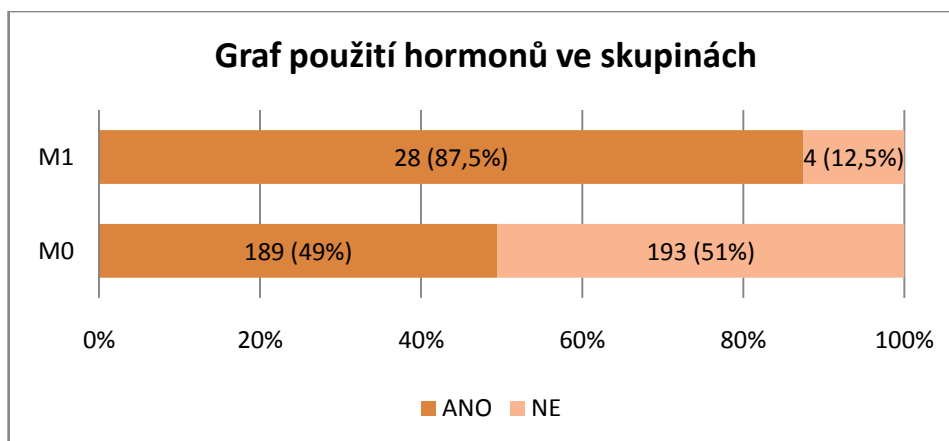
**H<sub>0</sub>:** Typ použité radioterapie nezávisí na typu operace.

**H<sub>A</sub>:** Typ použité radioterapie závisí na typu operace.

p-value =  $3,36 \cdot 10^{-144}$ , proto zamítáme nulovou hypotézu ve prospěch alternativní hypotézy  
**→NELZE TVRDIT, že typ použité radioterapie nezávisí na typu operace.**

#### 4. Kolik pacientů v jednotlivých stádiích mělo nějaký typ hormonální léčby

Ve skupině M0 je procentuelní rozložení pacientů, kteří podstoupili nějakou hormonální léčbu, srovnatelné s pacienty, kteří tuhle léčbu nepodstoupili. Ve skupině M1 však již převládají pacienti, kteří hormonální léčbu podstoupili.



Obrázek 16: Graf použití hormonů ve skupinách

hormony	M0	M1
ANO	189 (49%)	28 (87,5%)
NE	193 (51%)	4 (12,5%)

Určení závislosti proměnných pomocí kontingenční tabulky:

Kontingenční tabulka závislosti podání hormonů na rizikové skupině:

Hormony	M0	M1	Celkový součet
NE	193	4	197
ANO	189	28	217
Celkový součet	382	32	414

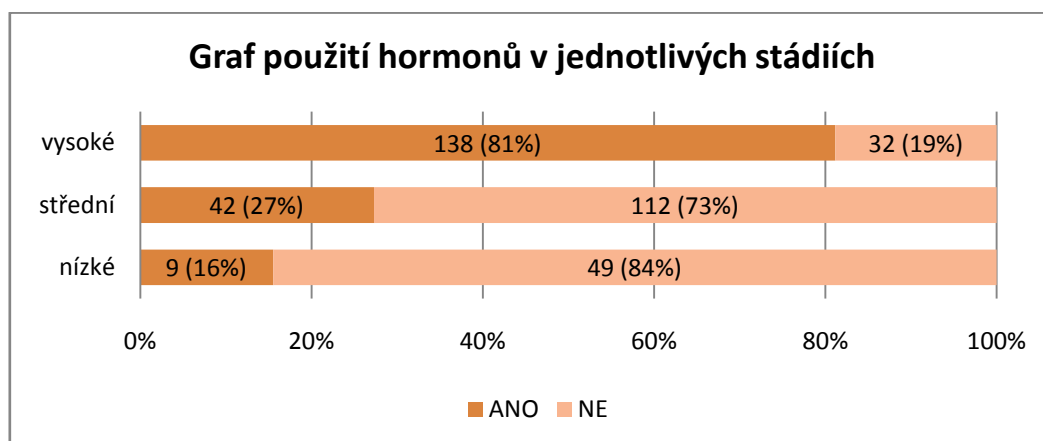
Tabulka 5: Závislost podání hormonů na rizikové skupině

Volba nulové a alternativní hypotézy:

**H0:** Podání hormonů pacientům nezávisí na rizikové skupině.

**HA:** Podání hormonů pacientům závisí na rizikové skupině.

p-value =  $3,52 \cdot 10^{-5}$ , proto zamítáme nulovou hypotézu ve prospěch alternativní hypotézy → **NELZE TVRDIT, že podání hormonů pacientům nezávisí na rizikové skupině.**



Obrázek 17: Graf použití hormonů v jednotlivých stádiích

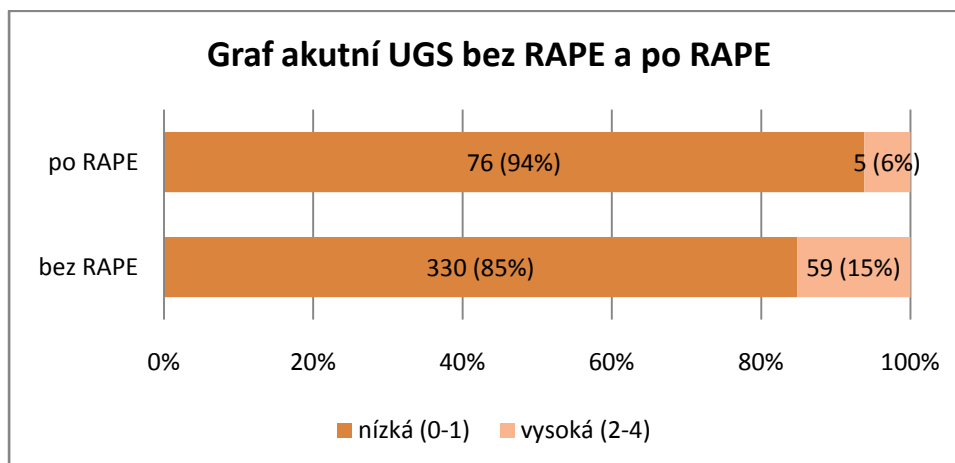
hormony	nízké	střední	vysoké
<b>ANO</b>	9 (16%)	42 (27%)	138 (81%)
<b>NE</b>	49 (84%)	112 (73%)	32 (19%)

Pokud se podíváme na pacienty v jednotlivých stádiích skupiny M0, vidíme, že v nízkém a středním stádiu převládají pacienti, kteří hormonální léčbu nepodstoupili, zatímco ve vysokém stádiu je tomu naopak a výrazně převládají pacienti s hormonální léčbou.



## 5. Jaké je procentuální zastoupení jednotlivých stupňů akutní a pozdní toxicity (akutní RTOG a pozdní RTOG) po RAPE a bez RAPE

Následující čtyři grafy týkající se akutní a pozdní UGS a GIT se od sebe liší hlavně v tom, že akutní UGS a GIT se týká většího počtu pacientů, avšak co se týče počtu pacientů, kteří trpí akutní popřípadě pozdní UGS nebo GIT, vidíme, že je vždy větší počet pacientů, kteří RAPE nepodstoupili a z nich převažují ti, kteří mají nízkou toxicitu. U pacientů, kteří RAPE podstoupili, také převažuje nízká akutní i pozdní toxicita.



Obrázek 18: Graf akutní UGS bez RAPE a po RAPE

aUGS	bez RAPE	po RAPE
nízká (0-1)	330 (85%)	76 (94%)
vysoká (2-4)	59 (15%)	5 (6%)

**Určení závislosti proměnných pomocí kontingenční tabulky:**

Kontingenční tabulka pro závislost akutní UGS na podstoupení RAPE:

Akutní UGS	0	1	Celkový součet
nízká	330	76	406
vysoká	59	5	64
Celkový součet	389	81	470

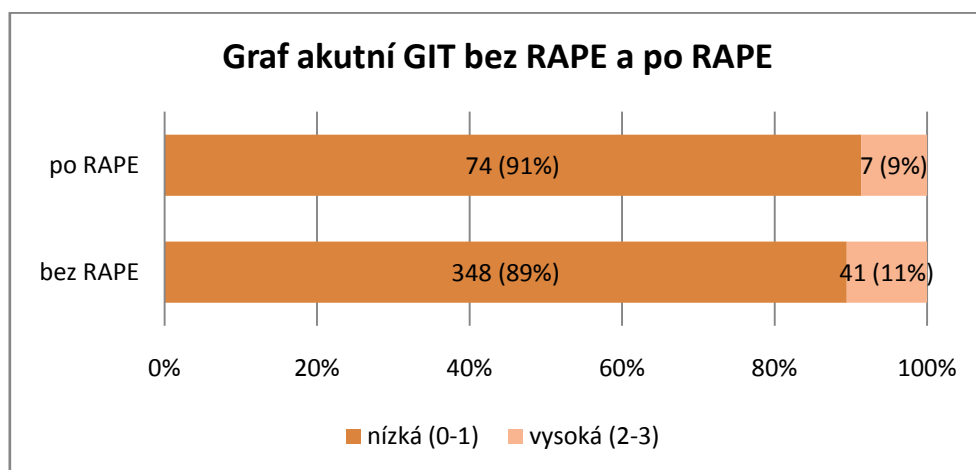
Tabulka 6: Závislost akutní UGS na podstoupení RAPE

Volba nulové a alternativní hypotézy:

**H<sub>0</sub>:** Stupeň toxicity akutní UGS nezávisí na tom, zda pacient podstoupil RAPE.

**H<sub>A</sub>:** Stupeň toxicity akutní UGS závisí na tom, zda pacient podstoupil RAPE.

p-value = 0,03, proto zamítáme nulovou hypotézu ve prospěch alternativní hypotézy → **NELZE TVRDIT, že stupeň toxicity akutní UGS nezávisí na tom, zda pacient podstoupil RAPE.**



Obrázek 19: Graf akutní GIT bez RAPE a po RAPE

aGIT	bez RAPE	po RAPE
nízká (0-1)	348 (89%)	74 (91%)
vysoká (2-3)	41 (11%)	7 (9%)

**Určení závislosti proměnných pomocí kontingenční tabulky:**

Kontingenční tabulka pro závislost akutní GIT na podstoupení RAPE:

Akutní GIT	0	1	Celkový součet
nízká	348	74	422
vysoká	41	7	48
Celkový součet	389	81	470

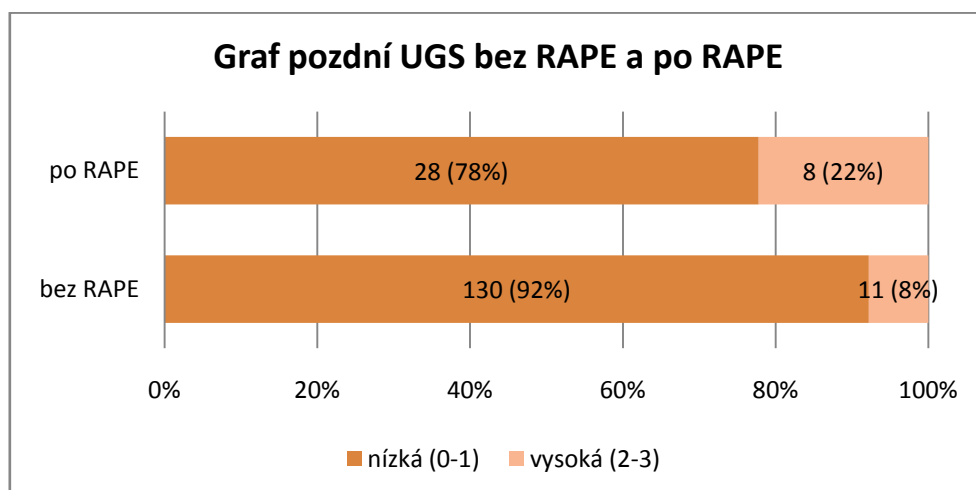
Tabulka 7: Závislost akutní GIT na podstoupení RAPE

Volba nulové a alternativní hypotézy:

H<sub>0</sub>: Stupeň toxicity akutní GIT nezávisí na tom, zda pacient podstoupil RAPE.

H<sub>A</sub>: Stupeň toxicity akutní GIT závisí na tom, zda pacient podstoupil RAPE.

p-value = 0,6, proto nezamítáme nulovou hypotézu → **Stupeň toxicity akutní GIT nezávisí na tom, zda pacient podstoupil RAPE.**



Obrázek 20: Graf pozdní UGS bez RAPE a po RAPE

pUGS	bez RAPE	po RAPE
nizká (0-1)	130 (92%)	28 (78%)
vysoká (2-3)	11 (8%)	8 (22%)

**Určení závislosti proměnných pomocí kontingenční tabulky:**

Kontingenční tabulka pro závislost pozdní UGS na podstoupení RAPE:

Pozdní UGS	0	1	Celkový součet
nizká	130	28	158
vysoká	11	8	19
Celkový součet	141	36	177

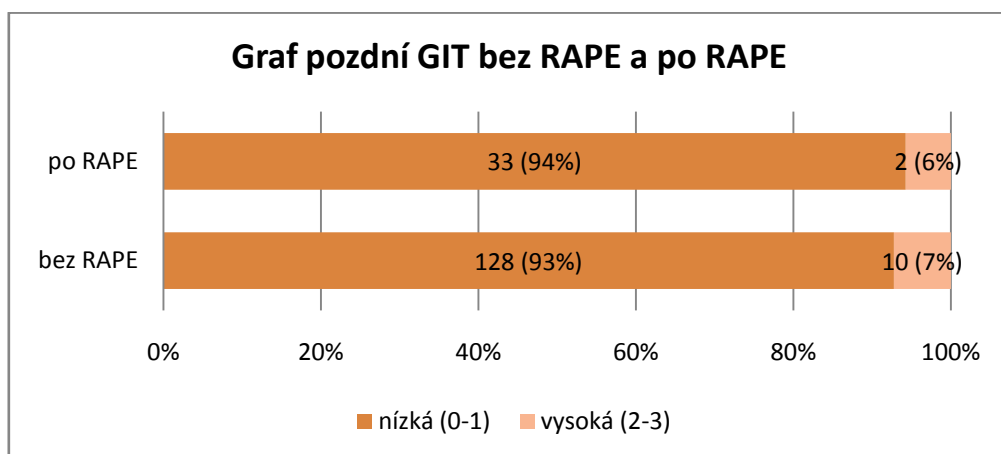
Tabulka 8: Závislost pozdní UGS na podstoupení RAPE

Volba nulové a alternativní hypotézy:

**H<sub>0</sub>:** Stupeň toxicity pozdní UGS nezávisí na tom, zda pacient podstoupil RAPE.

**H<sub>A</sub>:** Stupeň toxicity pozdní UGS závisí na tom, zda pacient podstoupil RAPE.

p-value = 0,01, proto zamítáme nulovou hypotézu ve prospěch alternativní hypotézy → **NELZE TVRDIT, že stupeň toxicity pozdní UGS nezávisí na tom, zda pacient podstoupil RAPE.**



Obrázek 21: Graf pozdní GIT bez RAPE a po RAPE

pGIT	bez RAPE	po RAPE
nizká (0-1)	128 (93%)	33 (94%)
vysoká (2-3)	10 (7%)	2 (6%)

**Určení závislosti proměnných pomocí kontingenční tabulky:**

Kontingenční tabulka pro závislost pozdní GIT na podstoupení RAPE:

Pozdní GIT	0	1	Celkový součet
nizká	128	33	161
vysoká	10	2	12
Celkový součet	138	35	173

Tabulka 9: Závislost pozdní GIT na podstoupení RAPE

Volba nulové a alternativní hypotézy:

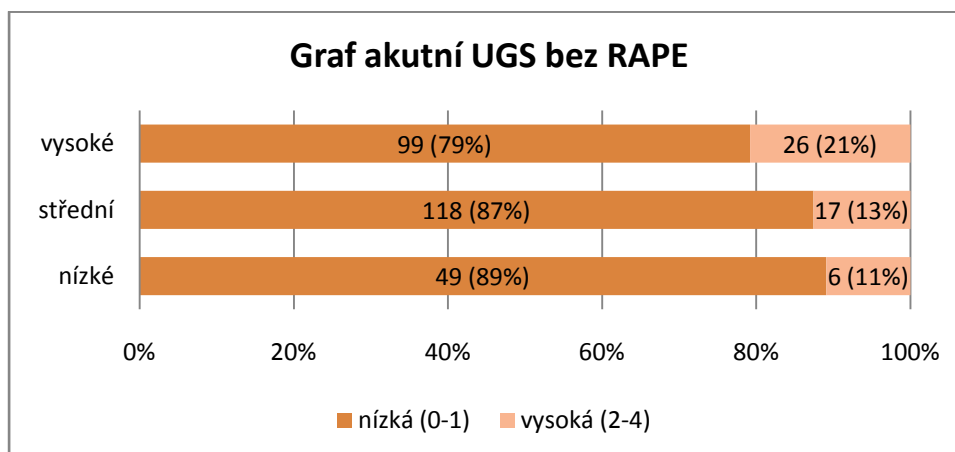
**H<sub>0</sub>:** Stupeň toxicity pozdní GIT nezávisí na tom, zda pacient podstoupil RAPE.

**H<sub>A</sub>:** Stupeň toxicity pozdní GIT závisí na tom, zda pacient podstoupil RAPE.

p-value = 0,75, proto nezamítáme nulovou hypotézu → **Stupeň toxicity pozdní GIT nezávisí na tom, zda pacient podstoupil RAPE.**

Následují grafy, zobrazující akutní a pozdní UGS a GIT u pacientů v jednotlivých stádiích skupiny M0, kteří RAPE nepodstoupili.

U obou akutních toxicit je nejvíce zastoupena ve středním a vysokém stádiu nízká toxicita. Vysoká toxicita se více objevuje u akutní UGS. U obou akutních toxicit se vysoká toxicita objevuje nejvíce ve vysokém stádiu.



Obrázek 22: Graf akutní UGS bez RAPE

aUGS	nízké	střední	vysoké
nízká (0-1)	49 (89%)	118 (87%)	99 (79%)
vysoká (2-4)	6 (11%)	17 (13%)	26 (21%)

**Určení závislosti proměnných pomocí kontingenční tabulky:**

Kontingenční tabulka pro závislost akutní UGS bez RAPE na stádiích M0:

Akutní UGS bez RAPE	nízké	střední	vysoké	Celkový součet
nízká	49	118	98	265
vysoká	6	17	27	50
Celkový součet	55	135	125	315

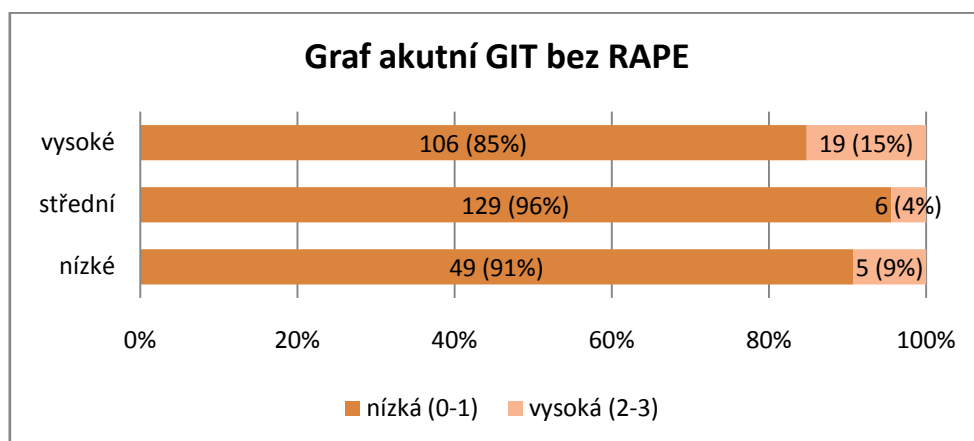
Tabulka 10: Závislost akutní UGS bez RAPE na stádiích M0

Volba nulové a alternativní hypotézy:

**H0:** Akutní UGS bez RAPE nezávisí na stádiích M0.

**HA:** Akutní UGS bez RAPE závisí na stádiích M0.

p-value = 0,075, proto nezamítáme nulovou hypotézu → **Akutní UGS bez RAPE nezávisí na stádiích M0.**



Obrázek 23: Graf akutní GIT bez RAPE

aGIT	nízké	střední	vysoké
nízká (0-1)	49 (91%)	129 (96%)	106 (85%)
vysoká (2-3)	5 (9%)	6 (4%)	19 (15%)

**Určení závislosti proměnných pomocí kontingenční tabulky:**

Kontingenční tabulka pro závislost akutní GIT bez RAPE na stádiích M0:

Akutní GIT bez RAPE	nízké	střední	vysoké	Celkový součet
nízká	50	129	105	284
vysoká	5	6	20	31
Celkový součet	55	135	125	315

Tabulka 11: Závislost akutní GIT bez RAPE na stádiích M0

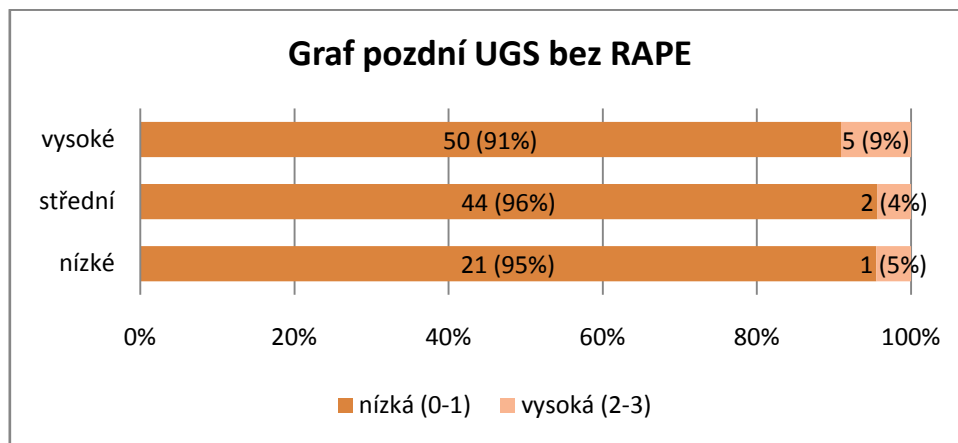
Volba nulové a alternativní hypotézy:

**H0:** Akutní GIT bez RAPE nezávisí na stádiích M0.

**HA:** Akutní GIT bez RAPE závisí na stádiích M0.

p-value = 0,007, proto zamítáme nulovou hypotézu ve prospěch alternativní → **NELZE TVRDIT, že akutní GIT bez RAPE nezávisí na stádiích M0.**

Pozdní UGS a GIT se opět týká menšího počtu pacientů než toxicita akutní. U obou pozdních toxicit je ve všech stádiích četnější nízká toxicita. U pozdní UGS převládá vysoká toxicita ve vysokém stádiu, naproti tomu u pozdní GIT se vysoká toxicita nejvíce vyskytuje rovnoměrně ve středním a vysokém stádiu.



Obrázek 24: Graf pozdní UGS bez RAPE

pUGS	nízké	střední	vysoké
nízká (0-1)	21 (95%)	44 (96%)	50 (91%)
vysoká (2-3)	1 (5%)	2 (4%)	5 (9%)

**Určení závislosti proměnných pomocí kontingenční tabulky:**

Kontingenční tabulka pro závislost pozdní UGS bez RAPE na stádiích M0:

Pozdní UGS bez RAPE	nízké+střední	vysoké	Celkový součet
nízká	65	50	115
vysoká	3	5	8
Celkový součet	68	55	123

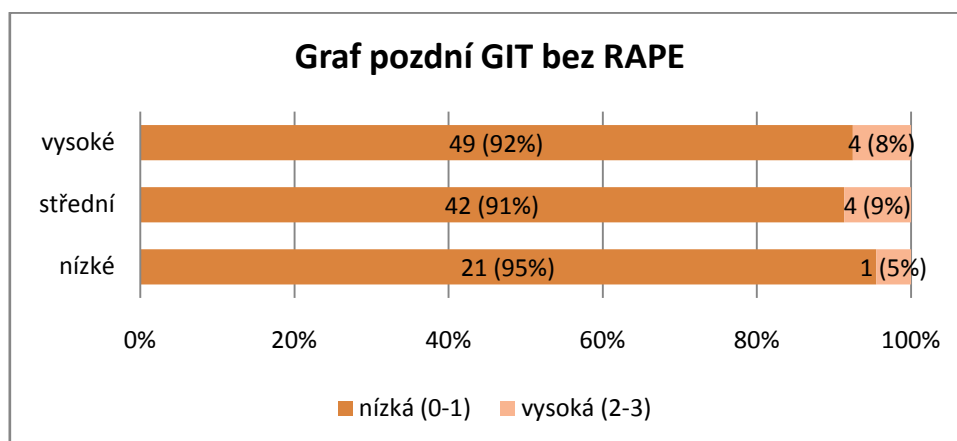
Tabulka 12: Závislost pozdní UGS bez RAPE na stádiích M0

Volba nulové a alternativní hypotézy:

**H0:** Pozdní UGS bez RAPE nezávisí na stádiích M0.

**HA:** Pozdní UGS bez RAPE závisí na stádiích M0.

p-value = 0,3, proto nezamítáme nulovou hypotézu → **Pozdní UGS bez RAPE nezávisí na stádiích M0.**



Obrázek 25: Graf pozdní GIT bez RAPE

pGIT	nízké	střední	vysoké
nízká (0-1)	21 (95%)	42 (91%)	49 (92%)
vysoká (2-3)	1 (5%)	4 (9%)	4 (8%)

**Určení závislosti proměnných pomocí kontingenční tabulky:**

Kontingenční tabulka pro závislost pozdní GIT bez RAPE na stádiích M0:

Pozdní GIT bez RAPE	nízké+střední	vysoké	Celkový součet
nízká	63	49	112
vysoká	5	4	9
Celkový součet	68	53	121

Tabulka 13: Závislost pozdní GIT bez RAPE na stádiích M0

Volba nulové a alternativní hypotézy:

**H0:** Pozdní GIT bez RAPE nezávisí na stádiích M0.

**HA:** Pozdní GIT bez RAPE závisí na stádiích M0.

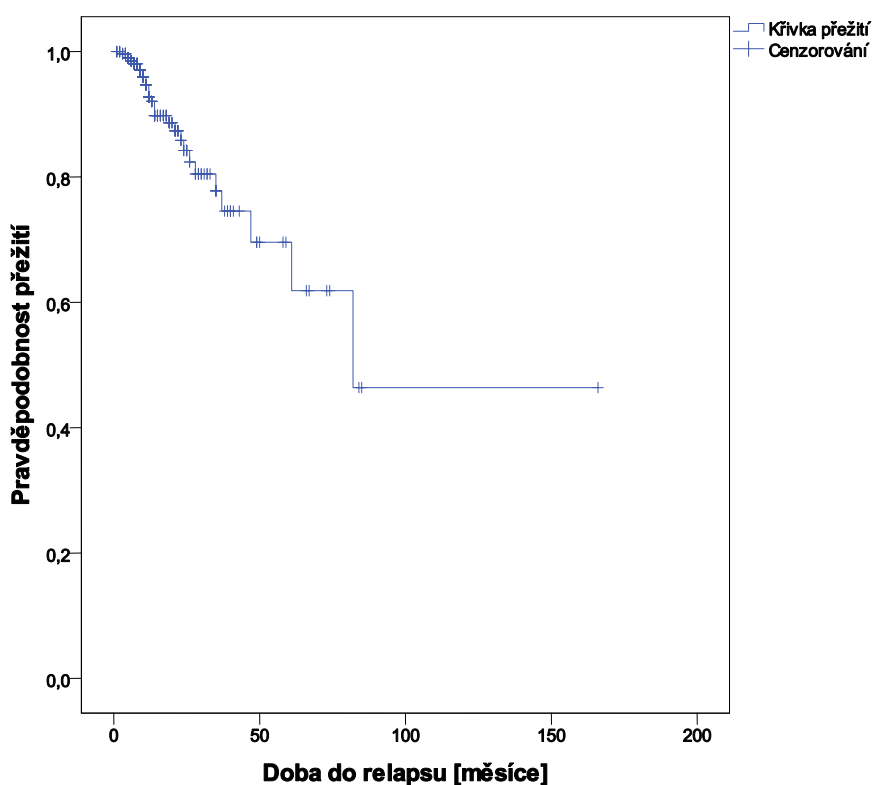
p-value = 0,97, proto nezamítáme nulovou hypotézu → **Pozdní GIT bez RAPE nezávisí na stádiích M0.**



## 6. Zpracování křivek přežití

V této části se budeme zabývat křivkami přežití pacientů, které byly vyhotoveny statistickým softwarem SPSS 18.0. Všechny použité data ke křivkám jsou v příložené příloze [III.] a [IV.]. Podíváme se na některé rozdíly mezi jednotlivými křivkami pacientů spadajících do určitých skupin. Jelikož křivky přežití budou náplní mého učiva v magisterském studiu, podrobnější analýzou těchto křivek bych se chtěla věnovat dále v diplomové práci.

Nejdříve se podíváme na celkovou křivku přežití, která se týká 316ti pacientů, bez jakéhokoliv zaměření léčby.



Obrázek 26: Celková křivka přežití

Výstup ze softwaru SPSS:

Tabulka shrnutí dat

Celkové N	N v události	Cenzorováno	
		N	Procenta
316	29	287	90,8%

Tabulka 14: Shrnutí pro celkovou křivku přežití

Z celkově 316ti pacientů bylo 90,8% (287 pacientů) dat cenzorováno.

### Střední hodnoty pro dobu přežití

Střední hodnota <sup>a</sup>		
Odhad	95% interval spolehlivosti	
	Spodní mez	Horní mez
102,190	73,429	130,951

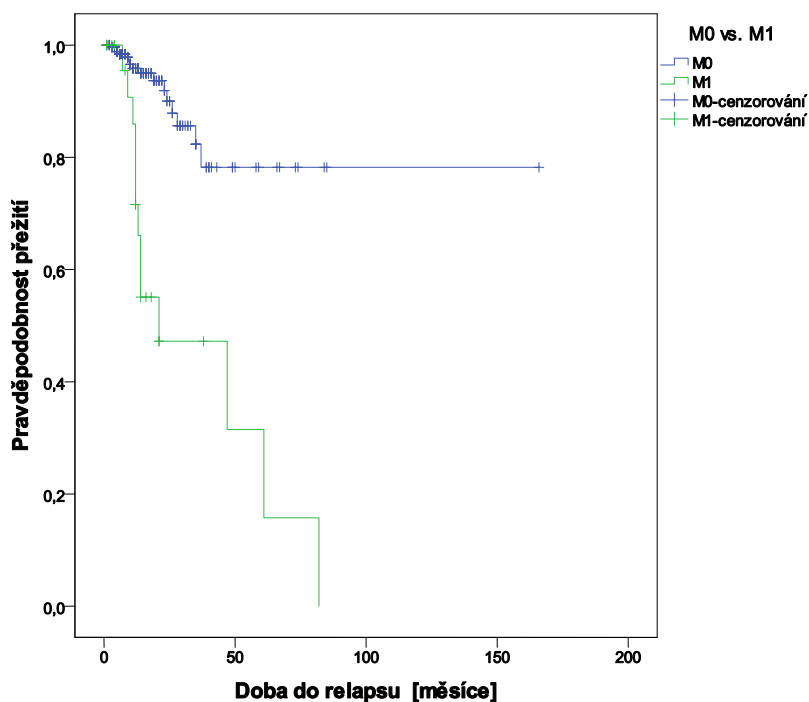
Tabulka 15: Tabulka středních hodnot pro celkovou křivku přežití

95%-ní interval spolehlivosti leží mezi hodnotami 73,429 a 130,951.

Percentily		
25,0%	50,0%	75,0%
Odhad	Odhad	Odhad
.	82,000	37,000

Tabulka 16: Tabulka percentilů pro celkovou křivku přežití

Po zhlédnutí křivek pro pacienty zvláště ve skupině M0 a M1 vidíme, že křivka M1 klesá mnohem strměji než křivka M0. Očekávaná doba do progresu u M0 je 40 měsíců. Skupina M0 zahrnuje 290 pacientů a skupina M1 zbylých 26 pacientů.



Obrázek 27: Křivky přežití pro M0 a M1

**Výstup ze softwaru SSPS:**

**Tabulka shrnutí dat**

M0 vs. M1	Celkové N	N v události	Cenzorováno	
			N	Procenta
M0	290	16	274	94,5%
M1	26	13	13	50,0%
Celkově	316	29	287	90,8%

Tabulka 17: Shrnutí křivek přežití pro M0 a M1

Z 316ti pacientů je ve skupině M0 cenzorovaných 94,5% (274) dat z 290ti pacientů a ve skupině M1 50% (13) dat z 26ti pacientů.

**Střední hodnoty pro dobu přežití**

M0 vs. M1	Střední hodnota <sup>a</sup>		
	Odhad	95% interval spolehlivosti	
		Spodní mez	Horní mez
M0	135,244	118,194	152,295
M1	36,796	22,197	51,395
Celkově	102,190	73,429	130,951

Tabulka 18: Tabulka středních hodnot křivek přežití pro M0 a M1

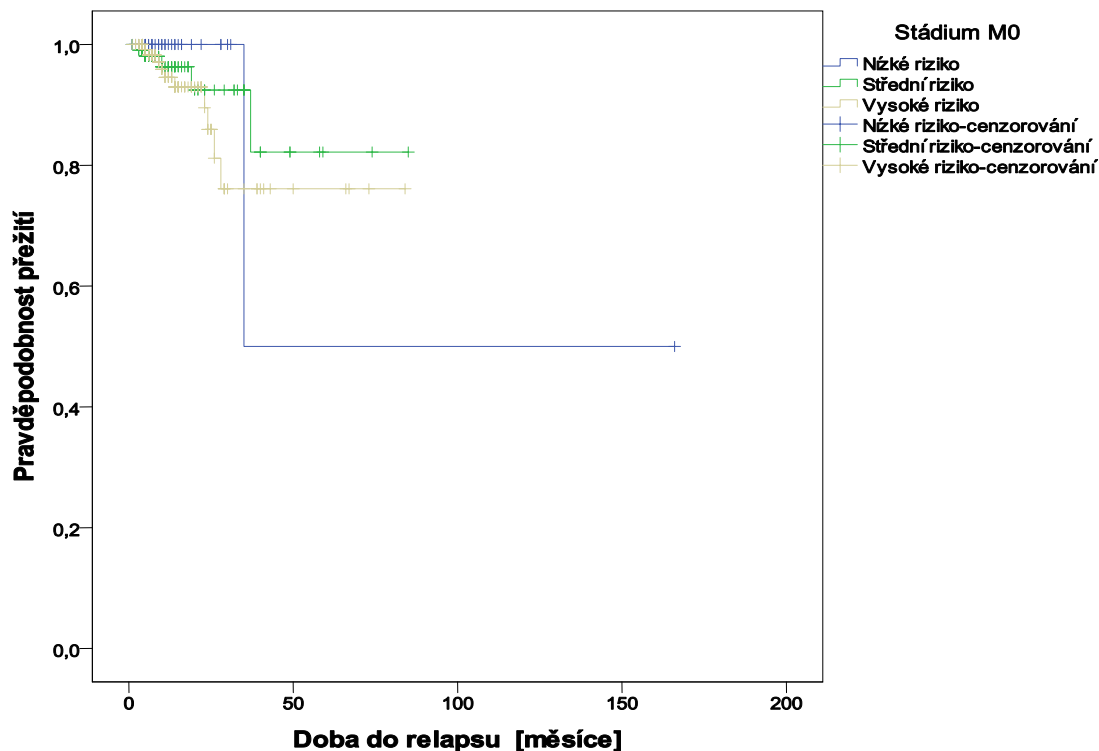
95%-ní interval spolehlivosti skupiny M0 leží mezi hodnotami 118,194 a 152,295 a u skupiny M1 je interval spolehlivosti od 22,197 do 51,395.

**Percentily**

M0 vs. M1	25,0%	50,0%	75,0%
	Odhad	Odhad	Odhad
M0	.	.	.
M1	61,000	21,000	12,000
Celkově	.	82,000	37,000

Tabulka 19: Tabulka percentilů křivek přežití pro M0 a M1

Dále se podíváme na křivky přežití pro jednotlivá stádia skupiny M0. Z grafu vidíme, že u pacientů v nízkém stádiu se relaps objevil pouze jednou, očekávaná doba do progresu je zde 35 měsíců. Nízké stádium obsahuje 38 pacientů, střední stádium 107 pacientů a vysoké stádium 145 pacientů.



Obrázek 28: Křivky přežití stádií M0

Výstup ze softwaru SSPS:

Tabulka shrnutí dat				
Stádium M0	Celkové N	N v události	Cenzorováno	
			N	Procenta
Nízké riziko	38	1	37	97,4%
Střední riziko	108	5	103	95,4%
Vysoké riziko	144	10	134	93,1%
Celkově	290	16	274	94,5%

Tabulka 20: Shrnutí křivek přežití pro stádia M0

V nízkém stádiu je cenzorovaných 97,4% (37) dat z 38 pacientů, ve středním stádiu 95,4% (103) dat ze 108 pacientů a ve vysokém stádiu 93,1% (134) dat ze 144 pacientů.

### Střední hodnoty pro dobu přežití

Stádium M0	Střední hodnota <sup>a</sup>		
	Odhad	95% interval spolehlivosti	
		Spodní mez	Horní mez
Nízké riziko	100,500	9,722	191,278
Střední riziko	74,652	64,069	85,235
Vysoké riziko	68,929	59,364	78,495
Celkově	135,244	118,194	152,295

Tabulka 21: Tabulka středních hodnot pro křivky přežití stádií M0

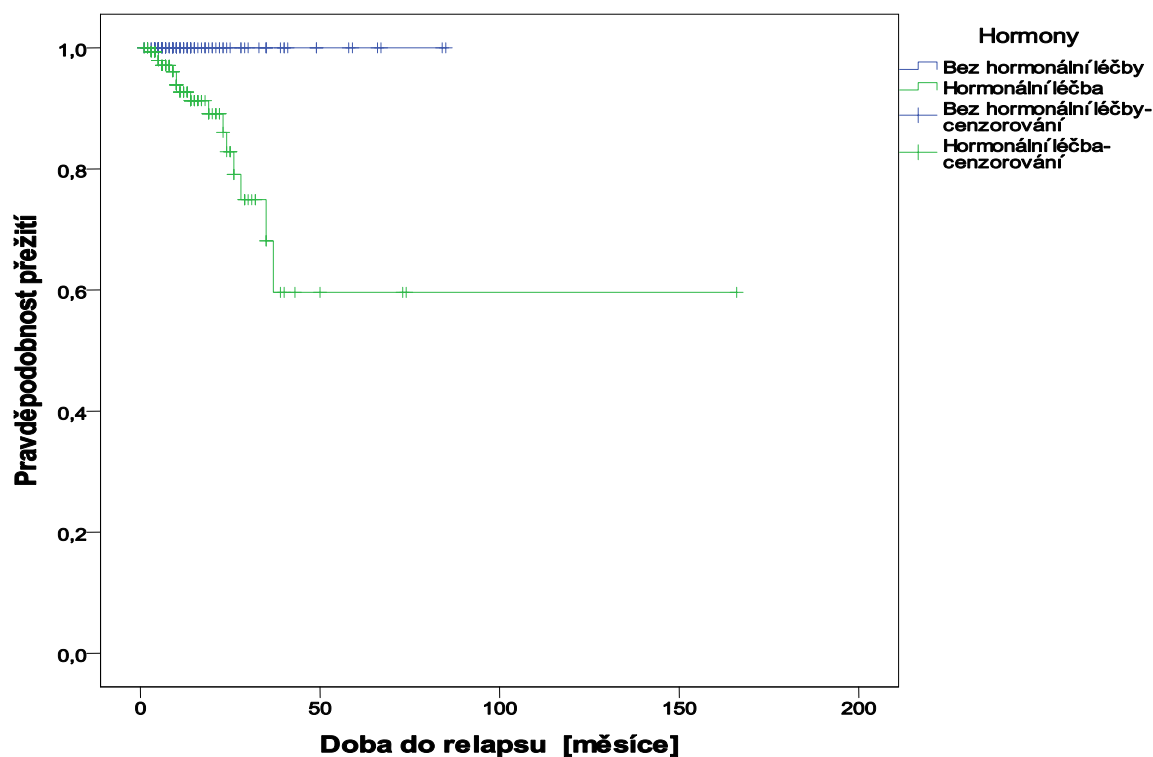
95%-ní interval spolehlivosti nízkého stádia leží mezi hodnotami 9,722 a 191,278, u nízkého stádia je interval spolehlivosti od 64,069 do 85,235 a u vysokého stádia spadá do intervalu 59,364 – 78,495.

### Percentily

M0 vs. M1	25,0%	50,0%	75,0%
	Odhad	Odhad	Odhad
M0	.	.	.
M1	61,000	21,000	12,000
Celkově	.	82,000	37,000

Tabulka 22: Tabulka percentilů křivek přežití pro stádia M0

A v neposlední řadě srovnáme křivky přežití pacientů ve skupině M0 zvlášť pro pacienty, kteří podstoupili hormonální léčbu, a kteří hormonální léčbu nepodstoupili. 129 pacientů hormonální léčbu nepodstoupilo, naproti tomu 161 pacientů léčbu podstoupilo.



Obrázek 29: Křivky přežití M0 se závislostí na hormonech

U pacientů, kteří hormonální léčbu nepodstoupili, je křivka přežití pouze konstantní přímkou, což znamená, že u žádného z pacientů nedošlo k relapsu. Naopak u pacientů, kteří hormonální léčbu podstoupili, je očekávaná doba do progresu asi 36 měsíců.

#### Výstup ze softwaru SSPS:

Tabulka shrnutí dat

Hormony	Celkové N	N v události	Cenzorováno	
			N	Procenta
Bez hormonální léčby	129	0	129	100,0%
Homonální léčba	161	16	145	90,1%
Celkově	290	16	274	94,5%

Tabulka 23: Shrnutí křivek přežití pro závislost skupiny M0 na hormonech

Ve skupině M0 u pacientů, kteří nepodstoupili hormonální léčbu, jsou všechna data 129ti pacientů cenzorovaná. U pacientů, kteří hormonální léčbu podstoupili, je cenzorováno 90,1% (145) dat ze 161 pacientů, kteří patří do této skupiny.

## 9. Závěr

Cílem mé bakalářské práce bylo zodpovědět otázky pro lékaře z onkologického oddělení ostravské Fakultní nemocnice. Jako efektivní nástroj pro zpracování jednotlivých otázek byly využity metody vyspělé statistické indukce. Pro vyšetřování závislostí mezi proměnnými byly použity testy nezávislosti v kontingenčních tabulkách. Jelikož zpracovávaná databáze obsahovala cenzorovaná data, byla na sestrojení křivek přežití aplikována Kaplan-Meierova metoda. Analýza ukázala, že např. hormonální léčba má velký vliv na sestrojení křivek přežití. V budoucnu bych na vše chtěla navázat v diplomové práci na magisterském studiu, která by se měla zabývat podrobnější analýzou křivek přežití uvedených v závěru této práce.

Pro vypracování bakalářské práce jsem ze statistických softwarů využila hlavně Statgraphics plus 5.1, SSPS 18.0 a MS Excel 2007. S každou z těchto aplikací jsem se díky tomuto projektu naučila mnohem lépe pracovat a využívat je v praxi na konkrétních medicínských datech. Při celkovém zpracování mé práce jsem si mohla vyzkoušet, jak se statistika využívá v praxi a co všechno je potřeba zajistit před samotným zpracováním dat. Za velký přínos považuji to, že již v této době jiný student VŠB-TU pracuje na novém vzhledu databáze, která v budoucnu usnadní práci samotným lékařům i statistikům.

# Literatura

1. Vitalion.cz. *Vitalion. Lepší informace, lepší zdraví.* [Online] [Citace: 23. 2 2010.] <http://nemoci.vitalion.cz/rakovina-prostaty/>.
2. Biobran.cz. *Biobran.* [Online] [Citace: 23. 2 2010.] <http://www.biobran.cz/rakovina-prostaty-BioBran.html> .
3. **BRISĚ, Radim a LITSCHMANNOVÁ, Martina.** *STATISTIKA I. pro kombinované studium.* OSTRAVA : Computer Press, 2004.
4. **HURT, Jan.** *Teorie spolehlivosti.* PRAHA : Státní pedagogické nakladatelství, 1984.
5. **ZVÁROVÁ, Jana a MALÝ, Marek.** *Statistické metody v epidemiologii.* PRAHA : Karolinum, 2003. 80-246-0765-4.



# Přílohy

- [I.] upravena\_databaze C61.xlsx
- [II.] resene\_otazky.pdf
- [III.] Krivky\_preziti.zip
- [IV.] SSPS.zip

Upravovaná lékařská databáze v Excelu  
Podrobně vyřešené otázky pro nemocnici  
Data křivek přežití ve Statgraphicsu  
Data křivek přežití v SSPS